主　　编　张景中　陈民众

执行主编　李尚志
编　　委　何书元　任宏硕　郑志明
　　　　　文志英　王长平　孟实华

信息技术使数学更有力量

在这一册里，我们将学习有关算法、统计和概率的初步知识.

用各种各样的数学知识解决实际问题和各学科的理论问题，许多情形需要算出结果，需要有一套可以具体操作的办法，这就要有算法. 算法是数学及其应用的重要组成部分，是计算科学的重要基础. 在现代信息技术飞速发展的今天，算法在科学技术和社会发展中发挥着越来越大的作用，并且日益融入社会生活的许多方面.

在中国古代数学中，蕴含了丰富的算法思想. 针对社会生活中出现的数学问题，分门别类，提出有效的机械化的解答方法，即算法，正是中国古代数学的特色. 以几何学为主体的西方古代数学，主要特色虽然是公理化方法而不突出算法，但也出现了用算法解决问题的著名范例，即求最大公约数的欧几里得算法. 由于现代信息技术的发展，算法的研究和应用已经渗透到数学的各个分支. 算法思想已经成为现代人应当具备的一种数学素养.

同学们已经学过的许多解决数学问题的方法，像加减乘除，解方程或方程组，数学表达式求值，解三角形等等，其实都是算法. 通过十年的学习，虽然还没有引进算法这个词，但可以初步体会和感受算法的

思想．在此基础之上，我们将结合对具体数学实例的分析，了解算法的概念和结构，体验算法的程序框图在解决问题中的作用；通过模仿、操作、探索，学习设计程序框图表达解决问题的过程，学习根据程序框图写出算法语句的基本方法；在模拟解决实际问题的过程中，体会算法的基本思想和有效性，发展有条理的思考和表达能力，提高逻辑思维的能力．

现代社会是信息化的社会，人们常常需要收集数据，从所获得的数据中提取有价值的信息，以作出合理的决策．如何从庞杂多变的社会万象中收集有用的数据？如何将大量的数据整理得井井有条？如何从整理过的数据中挖掘出宝贵的信息？这需要下一番去粗取精，去伪存真，由表及里，由此及彼的工夫；需要科学理论和方法的指导．统计正是研究如何合理的收集、整理和分析数据的学科，它可以为人们制定决策提供依据．

自然界和社会生活中，有些事情的发生和发展有着清楚的因果关系，但也有大量的现象表现出偶然性，即随机现象．尤其是在日常的社会生活中，随机现象几乎处处可见．人们收集到的许多数据，也不可避免地具有随机性．概率是研究随机现象规律的学科，它为我们认识客观世界提供了重要的思维模式和解决问题的方法，同时也为统计学的发展提供了理论基础．因此，统计与概率的基础知识已经成为现代社会公民的必备知识．

从小学到初中的数学课里，同学们已经知道一些

统计和概率的有关问题和方法．在这一阶段的学习中，我们将结合更多的实际问题情景，学习随机抽样、样本估计总体和线性回归的基本方法；体会用样本估计总体及其特征的思想；通过解决实际问题，较为系统地经历数据收集与处理的全过程，体会统计思维与确定性思维的的差异．我们还将结合具体的实例，学习概率的某些基本性质和简单的概率模型，加深对随机现象的理解．我们还将动手动脑，通过有趣的实验、计算机或计算器的模拟，估计简单随机事件发生的概率．

用了计算机，可以更便捷地处理统计数据，模拟随机现象，以及执行形形色色的算法．希望大家尽可能地结合本册中的内容，学习有关的信息技术的知识和操作．在实际操作中，能够更切实地感受算法、统计和概率的思想．

祝同学们在新的学期里学得好，玩得好！

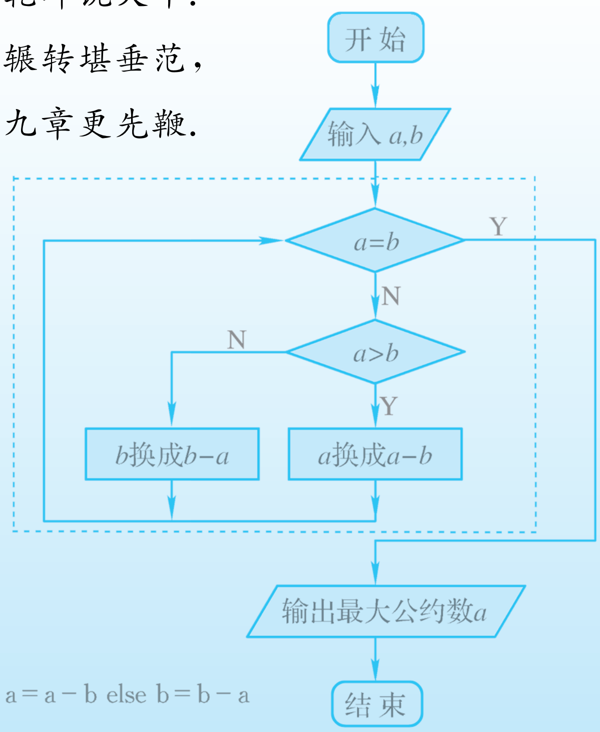
作　者

2006 年 12 月

目 录

第11章 算法初步

有规有矩成方圆，
步步为营释谜团.
亡羊歧路有判断，
戏马平川巧循环.
运筹帷幄操胜算，
袖里乾坤说大千.
欧公辗转堪垂范，
华夏九章更先鞭.



S1 Input a,b
S2 While a≠b
S3 If a>b then a=a−b else b=b−a
S4 End while
S5 Print a

我们学过很多解题方法，从加、减、乘、除到解方程和画函数图.

碰到一个题目，如果能够对号入座，归入学过的题型，便能依法炮制，手到擒来.

这就是说，对于这类题目，我们已经掌握了算法.

算法(algorithm)，就是对一类问题的有章可循的通用求解方法.

有了算法，就能让计算执行算法，解决问题.

一个算法，就是一个映射，也可以叫作一个函数. 问题中的已知数据就是函数的自变量，最后得到的答案就是函数值. 不过，这里的自变量可能不是一个数而是几个数，或是一些符号；函数值也可能是几个数，或一些符号.

这是一类特殊的函数：从自变量出发，能够一步一步地构造出函数值来. 所以，算法也可以看成是可构造的函数，或可计算的函数.

11.1 算法的概念和例子

用公式解题，当然有章可循，有一个公式也就有一种算法.

例如：一元二次方程 $x^2+2x-3=0$ 的根，可以用公式来计算：

$$x_1=\frac{-2+\sqrt{4+12}}{2}=1, \qquad x_2=\frac{-2-\sqrt{4+12}}{2}=-3.$$

但是，有些问题不一定能找到现成的公式解.

求两个正整数的最大公约数，就没有现成的公式.

虽然没有现成的公式，并非不能算.

例 1 求 273 和 147 的最大公约数.

解 由小到大地分解出 273 和 147 的公约数：

第一步，用公约数 3 分别除 273 和 147，得 91 和 49；

第二步，用公约数 7 分别除 91 和 49，得 13 和 7；

第三步，13 和 7 没有大于 1 的公约数，分解过程结束. 把已经分解出来的公约数 3 和 7 相乘，就得到 273 和 147 的最大公约数 21.

如果所给的两个数很大，上面的方法做起来就比较困难. 中国古代数学家创造了一种很容易执行的算法，叫“更相减损术”. 具体做法是：大数减小数，用差代替大数，重复进行，直到两数相等为止. 这个相等的数就是这两个数的最大公约数.

先做做，再想它的道理.

例 2 求 7 267 和 6 192 的最大公约数.

解 下面把全过程的步骤列出：

1	7 267	6 192
2	1 075（=7 267−6 192）	6 192
3	1 075	5 117（=6 192−1 075）
4	1 075	4 042（=5 117−1 075）
5	1 075	2 967（=4 042−1 075）
6	1 075	1 892（=2 967−1 075）
7	1 075	817（=1 892−1 075）
8	258（=1 075−817）	817
9	258	559（=817−258）
10	258	301（=559−258）
11	258	43（=301−258）
12	215（=258−43）	43
13	172（=215−43）	43
14	129（=172−43）	43
15	86（=129−43）	43
16	43（=86−43）	43

总共有 16 个步骤，最后一个步骤两边的数值相等，就是所要求的最大公约数.

这样做是因为，每一步左右两数虽然变了，但是两数的最大公约数却没有变. 即：

7 267 和 6 192 的最大公约数等于 7 267−6 192 和 6 192 的最大公约数……

到最后一步两数相等时，这个数自然就是最大公约数.

这样的求解过程，就是在执行求最大公约数的算法.

仔细分析上面这些步骤，可以看到还有改进的余地，从第 3 步到第 7 步，反复从右边减去同一个数 1 075，直到小于左边的数，其实这 5 个步骤改成除法只需一步就可以完成. 同样，从第 9 步到 11 步，反复 3 次从右边减去 258，3 步可并成一步. 从第 12 步到第 16 步，反复从左边减去 43，直到等于右边的数，这 5 个步骤改成除法也只

需一步就可以完成. 由此想到用除法代替减法，原来用 16 个步骤完成的工作，只需要 6 个步骤就能完成.

把“更相减损术”中的多次连减换成带余除法，就成了欧几里得在《几何原本》中叙述的“辗转相除法”. 但是欧几里得的“辗转相除法”比“更相减损术”要晚几百年.

上述方法条理清晰、步骤明确、可操作性强，是一种机械化的程式. 它体现出算法的特征.

习题 1

学而时习之

1. 写出后面这个四则题分步计算的次序.

$$a-b+c\times d\div(e+f)=?$$

2. 用“更相减损术”求 147 和 273 的最大公约数.

温故而知新

3. 用“辗转相除法”求 7 479 696 和 8 235 101 的最大公约数.（可利用计算器或计算机.）

4. 设计一个算法，求一个正整数的最小的不为 1 的约数.

11.2 算法的结构和程序框图

11.2.1 模拟计算机的游戏

要让计算机执行一个算法，必须根据算法编出计算机的程序.

我们已经知道了用“更相减损术”求最大公约数的算法.

但是，计算机如何执行这个算法呢？

让我们做一个用人来模拟计算机的游戏，从游戏中发现算法的结构，了解计算机如何工作.

游戏中有 6 个角色，分别由 6 位同学扮演. 6 个角色分别是判断器 1、判断器 2、执行器 A、执行器 B、输入器和输出器.

游戏的道具有：红蓝圆珠笔各一支，数据记录卡片一张. 所谓数据记录卡片，就是一张纸，中间画一条竖线分成 A，B 两栏；A 栏中的数据用红笔记录，B 栏中的数据用蓝笔记录.

游戏开始时，红蓝圆珠笔分别交给执行器 A，B；数据记录卡的 A，B 两栏的第一行分别用红蓝笔写下两个正整数 a 和 b，例如 a＝343，b＝277，再把记录卡交给输入器.

这 6 个角色的操作规则分别是：

S1 输入器：把写有数据 a，b 的记录卡交给判断器 1.

S2 判断器 1：每拿到记录卡时，检查新写下的红数字 a 和蓝数字 b 是否相等. 如果相等，就把记录卡给输出器，否则，就给判断器 2.

S3 判断器 2：拿到记录卡时，检查新写下的红数字 a 和蓝数字 b 哪个大. 如果红数字 a 大，就把记录卡给执行器 A，否则，就给执行器 B.

S4 执行器 A：拿到记录卡时，将当前的红色数字 a 减去蓝色数字 b，所得的差写在左栏的下一行，划掉上一行的数据，把记录卡交给判断器 1.

S5 执行器 B：拿到记录卡时，将当前的蓝色数字 b 减去红色数

字 a，所得的差写在 B 栏的下一行，划掉上一行的数据，把记录卡交给判断器 1.

S6 输出器：从判断器 1 手中接到记录卡时，宣读上面的计算结果，宣布计算任务完成.

只要每个角色按规定机械地操作，记录卡就会在几个人之间不停地传递，直到交给输出器．这时，记录卡上当前的红、蓝两数应当相等，它就是答案：a 和 b 的最大公约数.

如果全班分成几个小组来做，只要开始输入的两个数一样，各组最后得到的答案应当相同．如果不同，一定是某个“计算机”出了故障！

在模拟游戏过程中，除了输出器，每个角色完成一步操作后，总有一个角色接过记录卡，进行下一步操作.

有的角色，它的后继操作角色是确定的．例如，执行器 A 每次操作完成后总是把记录卡交给判断器 1．我们说，执行器 A 和判断器 1 两者组成一个“顺序结构”.

有的角色，它要根据某些条件来确定后继操作角色.

例如，判断器 1 手中的记录卡，根据 $a=b$ 是否成立，确定是交给判断器 2，还是交给输出器，我们说，由判断器 1、判断条件 $a=b$ 以及两个可能的后继操作角色组成一个“条件结构”，也叫“选择结构”或“分支结构”.

想一想，游戏过程中还有别的顺序结构和条件结构吗？

在游戏过程中，记录卡从判断器 1 手中传给判断器 2，接着传给执行器 A 或执行器 B，然后又回到判断器 1，经历了一个循环．这样的循环在计算过程中可能有多次，最后当条件 $a=b$ 成立时退出循环.

我们说，由判断器 1、判断器 2、执行器 A、执行器 B 以及退出循环的判断条件 $a=b$ 共同组成一个“循环结构”.

麻雀虽小，五脏俱全．在用“更相减损术”求最大公约数的算法中，包含了算法的三种基本结构：顺序结构、条件结构和循环结构.

下面的框图(图 11－1)画出了模拟游戏中的 6 个角色，以及数据记录卡在它们之间传递的路线．这样的框图，叫作算法程序框图.

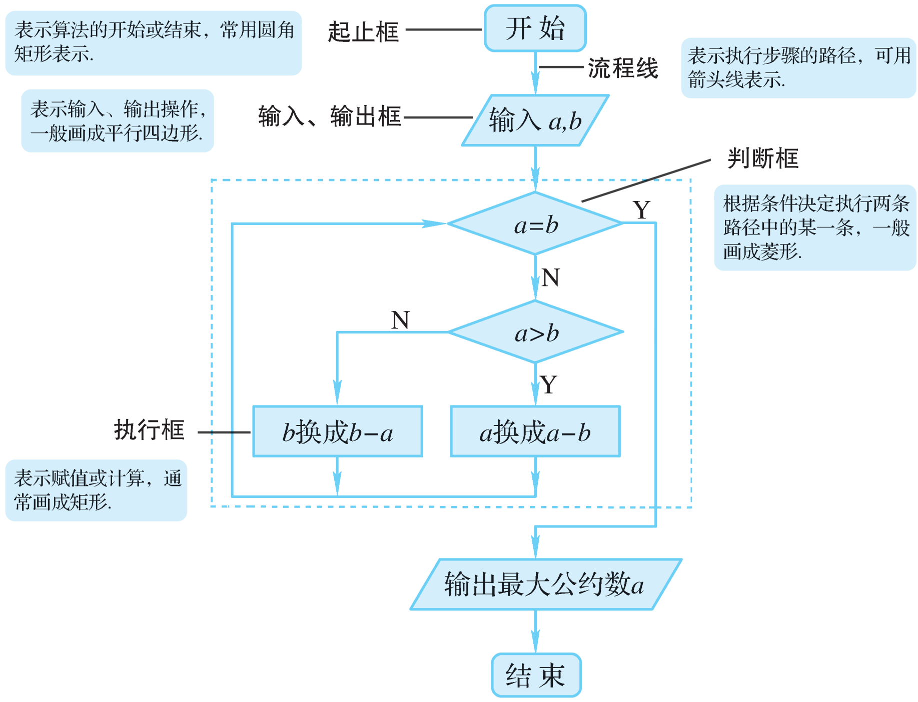


图 11－1

你能指出左边的算法程序框图中，有哪些顺序结构，有哪些条件结构和循环结构吗？

虚线框出的是循环体，每执行一次循环，两个数中至少有一个要变小，在有限次之后，总会使两数相等，此数即为最大公约数.

图中除了“开始”和“结束”，还有 6 个框，分别代表模拟游戏中的 6 个角色. 你能够指出哪个框代表哪个角色吗？

图中带箭头的线指明了游戏中记录卡传递的路线，也就是算法执行过程中操作的顺序. 你能说明从菱形框为何引出两条路线吗？两条路线旁边写的 N 和 Y 又是什么意思呢？

下面，用更多的例子，说明算法的这三种基本结构的应用.

课后思考和讨论

1. 在模拟游戏中，能不能只设一个判断器，由 5 个角色进行操作？
2. 如果用辗转相除法代替更相减损法，游戏的规则如何改变？算法的程序框图又该如何画呢？

11.2.2 顺序结构

有些事情，只要按顺序做，就能一气呵成.

例 1 鸡兔同笼问题.

笼子里有若干只鸡若干只兔. 鸡和兔共有 8 个头，22 只脚. 问笼子里有几只鸡，几只兔？

解 右下的程序框图(图 11－2)清楚地说明了解题过程.

该图是顺序结构，解题的过程是自然地由上而下依次执行，虽然是一种最简单的算法过程，但是它对所有的鸡兔同笼问题都适用.

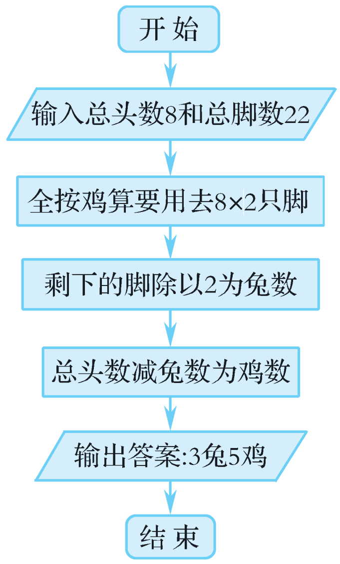


图 11－2

一个算法中如果没有条件结构和循环结构，只有顺序结构，在做模拟游戏时，除了输入器和输出器，所有的角色就都是执行器了.

你能设计一个求解一般的鸡兔同笼的模拟游戏吗？

如果游戏中的每个“执行器”只会一种运算，有的只会做乘法，有的只会做减法等，要安排几个执行器？

别看顺序结构简单，也能做不少问题.

如果有一个会根据公式做四则混合运算的“执行器”，要安排几个“执行器”？程序框图又是什么样呢？

例 2 已知平面上两个点 A，B，其距离小于 2，用一个只能画半径为 1 的圆的圆规，请画出第三个点 C，使得 $AB=BC=AC$.

解 如图 11－3：

这种圆规叫作“生锈的圆规”. 几何学家研究发现，这种圆规作图的本领还不小呢.

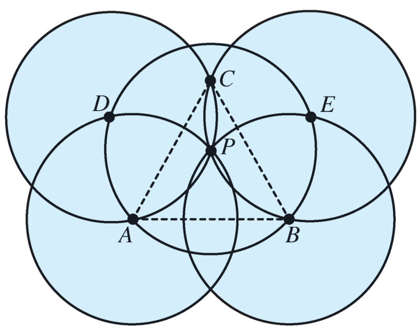


图 11－3

S1 分别以 A，B 为圆心作圆，在两圆交点中取一点 P；

S2 以 P 为圆心作圆，与两圆分别交于不同于 A，B 的点 D，E；

S3 分别以 D，E 为圆心作圆，交于不同于 P 的点于 C.

则 C 点就是所求的点.

几何学虽有 2 000 多年的历史，发现这个作图却是上世纪 70 年代的事. 可见结构很简单的算法，也会有丰富的内容.

练　习

1. 假定一个执行器只能做一种运算，设计一个求解一般的鸡兔同笼问题的模拟游戏.
2. 请写出用圆规直尺二等分一线段的流程图.

11.2.3　条件结构

在前面求最大公约数的模拟游戏中，条件结构起了关键作用.

比如，没有判断器 1 把关，就不知道何时退出循环，输出结果.

许多事情的处理过程中，即使没有循环，也可能有些不确定的因素，必须根据情况作出判断选择，以保证目标的实现. 判断器好比现场指挥，是见机行事的决策者.

例 1　北京的汪先生要参加明天在上海召开的会议，最好买今晚的末班机票赴会；如果买不到机票，就坐今晚的火车走；如果连火车票都买不到，就只好乘明天的首发航班去（假定次日的票总能保证）. 用框图表示他的购票方案.

解　汪先生的购票方案可以表示为如 P. 10 上面的框图(图11-4).

依此框图，无须多言，就知道该怎么处理事情.

即使用公式解题，有时也要判断一下题目中的已知数是否符合公式要求的条件，不要贸然行事，以免徒劳无功.

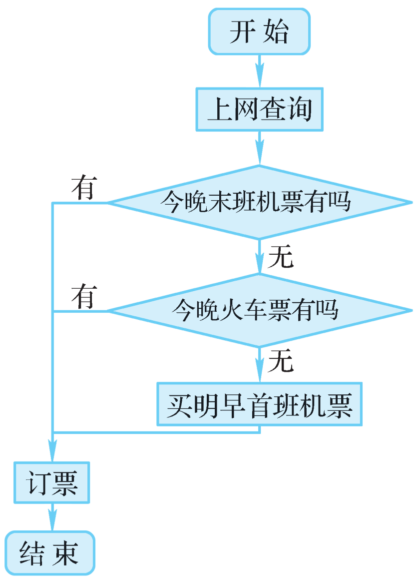


图 11－4

前面提到用公式求二次方程的根的问题，一般就要先计算一下方程的判别式 $\Delta=b^2-4ac$；如果 $\Delta<0$，则方程无实根，不要继续算了．

有时，情形变化较多，一个判断器还不够用．

例 2 给定三条线段，其长度分别是 a，b，c．画出判断此三线段能否成为一个三角形的三边的算法的程序框图．

解 算法原理：验证是否满足“三角形任意两边之和大于第三边”．如果满足条件，则能画三角形，否则不能．

一种算法的程序框图是：

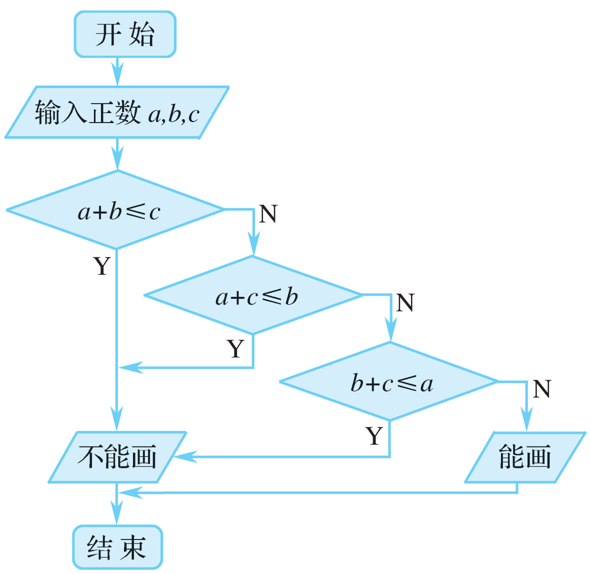


图 11－5

对于给定的条件，经判断后只有两种可能："是(Y)"或"否(N)". 对于有多种可能的情形，就可以像上图一样，让一个菱形图框接着一个菱形图框地顺排而下，既穷举了所有的可能性，又有条不紊地安排了整个工作的流程.

练习

1. 输入整数 a，按照下面的规则放置：

   如果除以 4 后，其余数等于 0，则放入红盒中；其余数等于 1，则放入黄盒中；其余数等于 2，则放入蓝盒中；其余数等于 3，则放入白盒中.

   画出放置数据的程序框图.
2. 画出计算符号函数 $\operatorname{sgn}(x)$ 的算法程序框图.
3. 画出用公式法解一元二次方程的程序框图.

11.2.4 循环结构

求最大公约数的模拟游戏中，还用到了循环结构.

循环结构做的是大量重复的同类型的操作，正好发挥出计算机的特长.

循环必须有始有终，何时退出循环，就要根据条件作判断. 所以，循环结构离不开条件结构.

求最大公约数的算法中，并不知道要循环多少次，只知道 $a=b$ 时退出循环. 但也有许多问题的算法，循环次数可以预先设定.

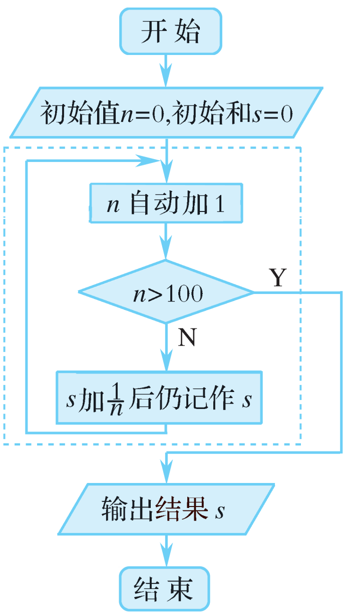


图 11－6

例 1 计算 $1+1/2+1/3+\cdots+1/100=?$

解 通过模拟游戏设计的思考，容易得出如右的程序框图(图 11－6).

图中有两个执行器，一个负责记录循环次数，每次加 1；一个负责计算所求的和，每次加 $\frac{1}{n}$. 一个判断器监视循环次数，够了就退出.

虽然每加一项要有三个动作，一个计数动作，一个判断动作，一个相加动作，但是这个虚框可以反复使用 100 次，比只用顺序结构来描述要合算得多.

例 2 用 1 000 块大小相同的正方体的积木，垒一个金字塔，自上而下第 n 层用 n^2 块积木，问最多能垒几层？画出一个算法的程序框图.

解 从 1 000 中顺次减去 1，4，9，16，…，记下减的次数，减到不够减时，所记下的次数就是答案. 程序框图如图 11－7.

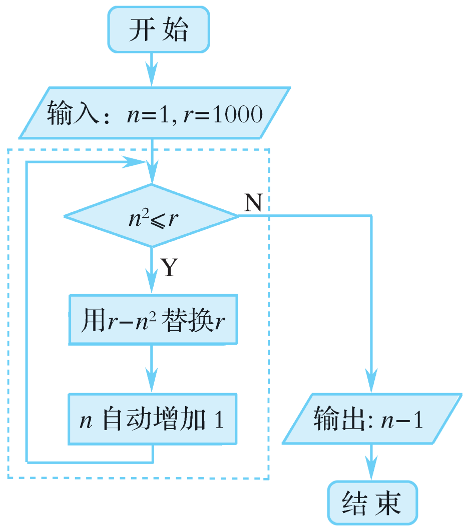


图 11－7

想一想，例 1 和例 2 的程序框图区别在哪里？

练习

1. 画出用累加法求小于 1 000 的正奇数的平方和的算法程序框图.
2. 让整数 n 从 1 开始增加，比较 1 $000n^3$ 和 2^n，问 n 等于多少时，2^n 第一次大于 1 $000n^3$？画出用逐项比较的方法解决这个问题的程序框图.

习题 2

学而时习之

1. 画出你真实经历过的一次看病的流程图.
2. 给定一个角$\angle AOB$，请写出二等分角的作图过程.
3. 写出解方程 $2x-7=0$ 的算法程序框图.
4. 根据税法，公民月薪所得不超过 800 元的部分不纳税，超过 800 元的部分为全月纳税所得额，适用超额累进税率，税款按下表分段累进计算.

全月应纳税所得额	税率
不超过 500 元的部分（含 500 元）	5%
超过 500 元至 2 000 元的部分（含 2 000 元）	10%
超过 2 000 元至 5 000 元的部分（含 5 000 元）	15%
超过 5 000 元至 10 000 元的部分（含 10 000 元）	20%

请用程序框图表示根据月工资（不超过 10 000 元）计算每月税后工资的算法.

温故而知新

5. 对于方程 $ax+b=0$，请解读下面这个程序在完成什么任务.

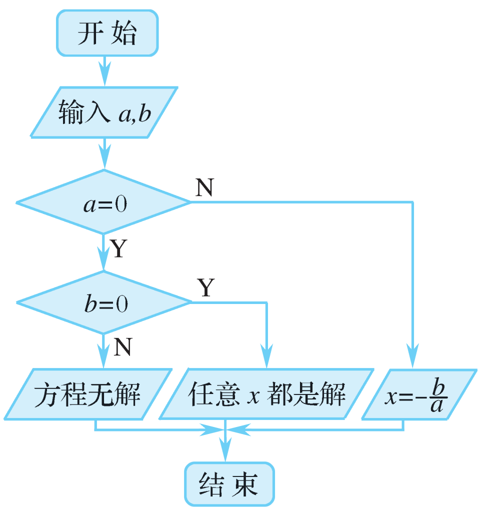


图 11－8

6. 如图 11－9，甲乙两人轮流剪珠子，一次剪一段，甲先剪，轮到谁没有珠子可剪，谁就是输家．请为甲设计一个取胜的程序框图．

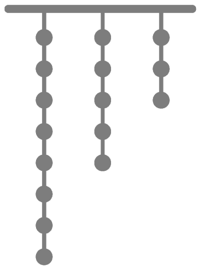


图 11－9

7. 一位商人有 9 枚银元，其中有 1 枚略轻的．请你设计一种算法，只用天平（不用砝码）将略轻的找出来．

8. 执行下面这个程序框图表示的算法会画出什么样的图形？

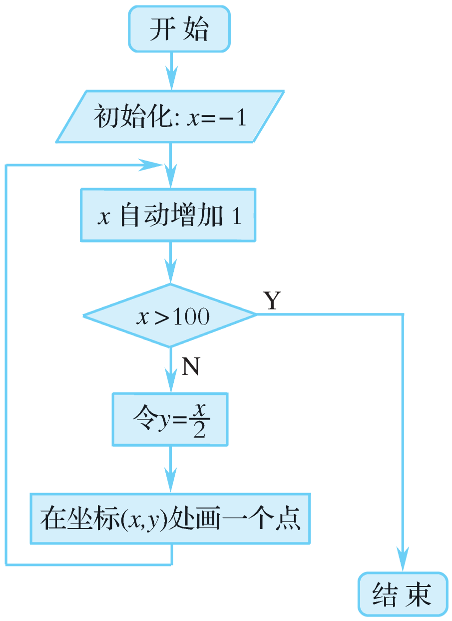


图 11－10

11.3 基本的算法语句

上面我们用自然语言或程序框图描述的算法，计算机是读不懂的，必须用计算机能理解的语言（代码）来表达，也就是编成计算机程序.

计算机能理解的语言种类很多，怎么办呢?

办法是先把算法表示成一种“算法语言”(也叫伪代码). 算法语言比较直观易懂，又简洁严谨，容易改写成各种语言的计算机程序. 下面用的算法语言接近于常用的 BASIC 语言.

11.3.1 把算法的描述变成伪代码

在求最大公约数的计算机模拟游戏(程序框图)中，对每个角色(图框)的任务和数据记录卡的传递方向做了详细的描述. 把这些描述逐条地“翻译”成算法语言，就得到了算法的伪代码.

S1 输入器：把写有数据 a，b 的记录卡交给判断器 1.

算法语言：Input a，b

S2 判断器 1：每拿到记录卡时，检查新写下的红数字 a 和蓝数字 b 是否相等. 如果相等，就把记录卡交给输出器，否则，就交给判断器 2.

算法语言：If a=b then goto S6
Else goto S3

S3 判断器 2：拿到记录卡时，检查新写下的红数字 a 和蓝数字 b 哪个大. 若红数字 a 大，就把记录卡给执行器 A，否则给执行器 B.

算法语言：If a>b then goto S4
Else goto S5

S4 执行器 A：拿到记录卡时，将当前的红色数字 a 减去蓝色数

字 b，所得的差写在左栏的下一行，划掉上一行的数据，把记录卡交给判断器 1.

算法语言：　a←a－b

　　　　　Goto S2

S5 执行器 B：拿到记录卡时，将当前的蓝色数字 b 减去红色数字 a，所得的差写在右栏的下一行，划掉上一行的数据，把记录卡交给判断器 1.

算法语言：　b←b－a

　　　　　Goto S2

S6 输出器：从判断器 1 手中接到记录卡时，宣读上面的计算结果，宣布计算任务完成.

算法语言：　Print a

把逐句翻译得到的算法语句集中起来，得到简洁准确的伪代码：

```
S1   Input a, b
S2   If a=b then goto S6
              Else goto S3
S3   If a>b then goto S4
              Else goto S5
S4   a←a－b
     Goto S2
S5   b←b－a
     Goto S2
S6   Print a
```

对照原来的中文描述，这些算法语句的意义很清楚. 后面，还要对它们作更具体的分类介绍.

课后思考和讨论

上面的伪代码比起原来的中文描述来，已经很简洁了. 你能不能

加以改进，使之更简洁呢？例如：

（1）既然 S2 下面就是 S3，S2 中的 Else goto S3 是否多余呢？

（2）将 S3 中的 Goto S4 和 Goto S5 分别换成 a←a－b 和 b←b－a，并且后面补一句 Goto S2，行不行呢？

这样改写之后，算法的伪代码就是：

```
S1  Input a, b
S2  If a=b then goto S5
S3  If a>b then a←a−b else b←b−a
S4  Goto S2
S5  Print a
```

这样仍能保持算法的原意，但更为简明.

由此可见，一个算法的伪代码可以有不同的写法. 只要能把算法说清楚，自己和别人都明白就行.

练　习

1. 把 11.2.4 中例 1 的算法用模拟游戏的方式描述，再翻译成伪代码.
2. 思考和讨论：能不能直接把程序框图翻译成伪代码？
3. 若有条件，将上面“课后思考和讨论”中的算法的 BASIC 程序在计算机上运行：

```
10 INPUT a, b
20 IF a=b GOTO 50
30 IF a>b THEN a=a−b ELSE b=b−a
40 GOTO 20
50 PRINT a
```

（程序运行开始，屏幕上出现问号，表示要你输入 a，b. 键入两个正整数，中间用逗号分开，再打 Enter 键，答案就出来了.）

11.3.2 输入输出语句与赋值语句

在上一节得到的伪代码由多行组成，每行都叫作一个“**算法语句**”，简称语句．第一个语句和最后一个语句分别是：

Input a，b

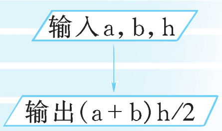
Print a

前者叫**输入语句**，意思是算法要求提供两个变量 a 和 b 的初始值以便继续工作；后者叫**输出语句**，意思是把变量 a 的当前值显示在屏幕上或打印出来．

一个输入语句可以输入一个或多个数据；一个输出语句也可以输出一个或多个数据．

算法程序框图中的输入输出框，在伪代码中对应于输入输出语句．

只用输入和输出语句，就能写出简单的算法的伪代码．



例 1 梯形两底 $a=3$，$b=5$，高 $h=4$，写出求梯形面积的算法的伪代码．

解 用一个输入语句和一个输出语句就可以了：

S1 Input a，b，h

S2 Print (a+b) * h/2

上述伪代码把语句编号 S1，S2 去掉，就可以在 BASIC 程序环境下运行．运行时屏幕上会出现问号提示你输入 3 个数．你用键盘或其他允许的方式输入 3，5 和 4 并打 Enter 键，输出的数据是梯形面积 16.

在上一节得到的伪代码中，还有 a←a−b 这样的语句．意思是用 a−b 的值代替 a 的当前值，也就是把 a−b 的值赋给变量 a．这样的语句叫作**赋值语句**．

算法程序框图中的执行框，在伪代码中主要表现为赋值语句．

用赋值语句把初始数据赋给变量，和输入语句有同样的功效．

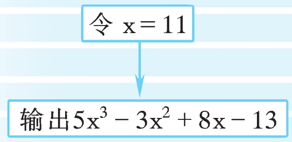
例 2 当 $x=11$ 时，写出求多项式 $f(x)=5x^3-3x^2+8x-13$ 的

值的算法的伪代码.

解 可以有两种写法:

算法一:

```
S1  x←11
S2  Print 5 * x^3−3 * x^2+8 * x−13
```



算法二:

```
S1  x←11
S2  Print ((5 * x−3) * x+8) * x−13
```

算法一和算法二的区别仅在于第二步中.

第一种算法用了 6 次乘法,第二种算法只用了 3 次乘法.因为四则运算中,加减法耗时最短,影响计算速度的主要因素来自乘法,所以后一种计算速度远比前一种要快.它是由中国古代数学家设计的,称为秦九韶算法.在大型计算中,迄今仍是通用的算法.

上述算法的伪代码,也可以多用一行赋值语句,写成:

```
S1  x←11
S2  f←((5 * x−3) * x+8) * x−13
S3  Print f
```

这样写的好处,是多保留了一个数据 f,可能在以后的其他计算中有用处.

练 习

1. 如图 11-11,这个程序运行后执行的是什么工作?
2. 写出计算 33,26,73 的平均值的算法的伪代码.
3. 华氏温度 h 和摄氏温度 c 的关系是

$$5(h-32)=9c.$$

写出由华氏温度 h 计算出摄氏温度 c 的算法的伪代码.

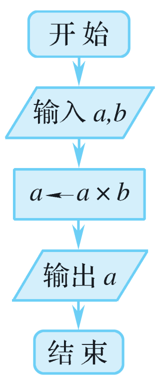


图 11-11

11.3.3 条件语句

在 11.3.1 节中得出的算法伪代码中，两次用到由 If 开头的语句，例如：

If a＝b then goto S6 else goto S3

这样的语句与程序框图中的判断框对应，叫作**条件语句**.

条件语句的一般形式为：

If A then B else C

其中 A 是对应的判断框里的条件，B 是条件成立时要执行的操作，C 是条件不成立时要执行的操作. 如果条件不成立时不需要做什么，后面也可以没有“else C”这个分支. 11.3.1 节在改进的算法伪代码中，语句 S2 就是不带 else 分支的条件语句.

如 11.2.3 中例 2 所示，条件结构中可能需要两个以上的判断器串联，对应的条件语句也会有两个或更多的 else 分支.

例 1 给定三条线段，其长度分别是 a，b，c. 写出判断此三线段能否成为一个三角形的三边的算法的伪代码（程序框图见图 11－5）.

解 对照程序框图，可写出伪代码如下：

S1 Input a，b，c

S2 If a＋b⩽c then print 不能成为三角形的三边

Else if a＋c⩽b then Print 不能成为三角形的三边

Else if b＋c⩽a then Print 不能成为三角形的三边

Else print 可以成为三角形的三边

End if

例 2 自来水公司对每户用水收费规定：每月用水量在 3 吨以内者，每吨收 1.1 元；每月用水量超过 3 吨且在 5 吨以内者，超过的部分，每吨收 1.6 元；每月用水量超过 5 吨者，超过 3 吨的部分，每吨收 2.2 元，请为自来水公司编写一个计费算法的伪代码.

解 下面就是所设计的算法程序框图和伪代码：

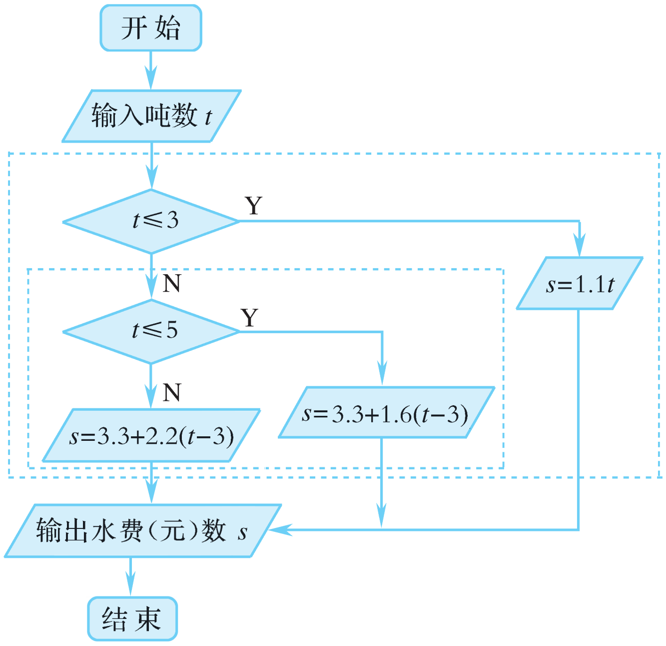


图 11－12

```
S1  Input t
S2  If t≤3 then s←1.1 * t
    Else if t≤5 then s←3.3+1.6 * (t−3)
    Else s←3.3+2.2 * (t−3)
    End if
S3  Print s
```

带有 Else if 分支的 if 语句，结束时要写上“End if”.

上面两例的条件语句都带有嵌套结构，即带有 Else if 分支，这样的条件语句，叫作复合条件语句.

练 习

1. 买一袋牛奶 2.8 元，买够 10 袋可打八折. 请设计一个根据所买袋数求出应付款数的计价算法，写出伪代码.
2. 写出求解一元二次方程 $x^2+cx+2=0$ 的算法的伪代码.
3. 设计一个从输入的 3 个数 a，b，c 中找出最大数的算法，画出程序框图并写出伪代码.

11.3.4 循环语句

我们看到，用输入、输出语句和赋值语句，可以实现顺序结构.

用条件语句，可以实现条件结构.

算法中的循环结构，是用什么语句实现的呢？

研究求最大公约数的算法的伪代码，就会发现，把条件语句和“goto”结合起来，就能实现循环，形如“Goto S2”的语句，叫转向语句，其意义自明，无需多谈.

此外，在常用的基本算法语句中，为了实现循环结构，还有专门的循环语句.

求最大公约数的算法中，事先不知道要循环多少次. 这时，可以使用“while”型循环语句，这种语句的一般形式是：

While A（条件）

B（若干语句）

End while

这叫“当型”循环，进入循环体之初，要在语句“While A”中判断，条件 A 成立时执行 B，否则退出或不进行循环，即跳过 B 和“End while”，执行下面的语句.

当执行到“End while”时，就自动返回到由 While 开始的地方，也可以把 B 用括弧括起来，就可以不写“End while”了.

从 While 到 End while 这些语句，构成一个循环体.

例 1 用“while”型语句，写出求最大公约数的算法伪代码.

解 S1 Input a，b

S2 While a≠b

S3 If a＞b then a＝a－b else b＝b－a

S4 End while

S5 Print a

如果事先知道循环的次数，可以用“For”型循环.

“For”语句的一般形式为：

For I（循环变量）from A（初值）to N（终值）step D（步长）

B（若干语句）

End for

意思是说：I 从 A 开始，每次加 D，到 N 为止，对每个 I 执行 B. 当步长 D=1 时，“step 1”可以省略；如果把 B 放在括弧里，“End for”可以省略.

从 For 到 End for 这些语句，构成一个循环体.

例 2 设计计算所有不超过 1 000 的正奇数的立方和的一个算法，使用“For”语句写出伪代码.

在计算机程序中，K^3 写成 K^3.

解 用“For”语句编写的伪代码如下：

```
S1 s←0
S2 For K from 1 to 999 step 2
S3 s←s+K^3
S4 End for
S5 Print s
```

在 BASIC 语言中，“End for”可用“Next B”代替，这里 B 是循环变量.

上述代码的第 2 行和第 3 行是说：K 从 1 开始，每次加 2，到 999 为止，对每个 K 执行 $s \leftarrow s+K^3$. K 叫循环变量.

编写循环语句，主要注意两点：一是入口的数据，二是出口的条件. 这也是程序最容易出错的地方.

练 习

1. 用“while”型语句，写出用辗转相除法求最大公约数的伪代码. 注意：a 除以 b 的余数用“a mod b”或 mod（a，b）表示.
2. 用“for”型语句，写出计算小于 100 的所有正整数的倒数和的算法的伪代码.
3. 不用循环语句，写出本节例 2 的伪代码.
4. 写出 11.2.4 节例 2 中算法的三种不同的伪代码.

多知道一点

“Until” 型循环语句

“While” 型循环语句着眼于何时进入和继续循环.

如果着眼于何时退出循环，可以用 “Until” 型循环语句，也叫 “直到” 型循环语句.

“Until” 型循环语句的一般形式为

Repeat

　　B（若干语句）

Until A（条件）

从 Repeat 开始，至少执行一次 B；每执行完 B 时判断条件 A 是否成立，成立则退出循环，否则又从 Repeat 开始.

从 Repeat 到 Until 这些语句，构成一个循环体.

求最大公约数的算法，用 “Until” 型循环语句可以写出下列的伪代码：

S1 Input a，b（a≠b）

S2 Repeat

S3 If a>b then a←a−b else b←b−a

S4 Until a=b

S5 Print a

这个算法要求输入的 a，b 不相等！为什么？

既然用 if 和 goto 就能够实现循环结构，为什么还要用专门的循环语句呢？

这是因为，由 goto 构成的循环语句，其循环体不容易看出，使用不当还会引起程序混乱，所以不提倡用 goto.

使用专用的循环语句，循环体可以一眼看出，方便多了.

习题 3

学而时习之

1. 写出计算三个数 a，b，c 的平均值的算法的伪代码.
2. 写出从四个数 w，x，y，z 中找出最小数的算法的伪代码.
3. 写出求 3 111，13 311 这两个数的最小公倍数的算法的伪代码.
4. 写出把分数 879/1 563 化成最简分数的算法的伪代码.
5. 写出判断一个正整数是否是素数的算法的伪代码.

温故而知新

6. 对于输入的三个正整数，下列伪代码描述的算法输出的是什么数？

```
S1 Input a，b，c
S2 Repeat
S3 If a>b then a←a−b
   Else If b>a then b←b−a
S4 If a>c then a←a−c
   Else If c>a then c←c−a
S5 If c>b then c←c−b
   Else If b>c then b←b−c
S6 Until a=b and b=c
S7 Print a
```

7. 使用算法基本语句写出将五个数字按大小排列的算法.
8. 某工厂有一批计时工，8 小时内每小时工资 6 元，8 小时外加班每小时 10 元，会计当天就要向工人清付工钱. 请编写一个根据工作小时数计算当天工资的算法伪代码.

11.4 算法案例

在科技活动和工程实践中，大量的计算工作使用着形形色色的算法，下面介绍的几个基本算法，问题虽然简单，其数学思想却引人入胜. 若能举一反三，反复求索，对提高数学素养和算法设计能力大有好处.

例 1 用二分法计算函数的零点.

在本套教材必修第一册的第 2 章 2.4.2 里，讲了可以用二分法计算函数 $f(x)$的零点，也就是求方程 $f(x)=0$ 的根. 现在进一步给出二分法的程序框图（图 11－13）和伪代码.

例如，求方程 $f(x)=x^5+x-1=0$ 在 $[0, 1]$ 上的根的近似值，精确到小数点后第三位，用二分法如何做呢？

函数 $f(x)$ 的值在$[0, 1]$两端异号，$[0, 1]$内必定有一根；把区间$[0, 1]$分成左右两部分，再检验这两个区间，哪个区间端点的函数值是异号的，方程的根必在此区间里，则留下这个区间再进行检验.

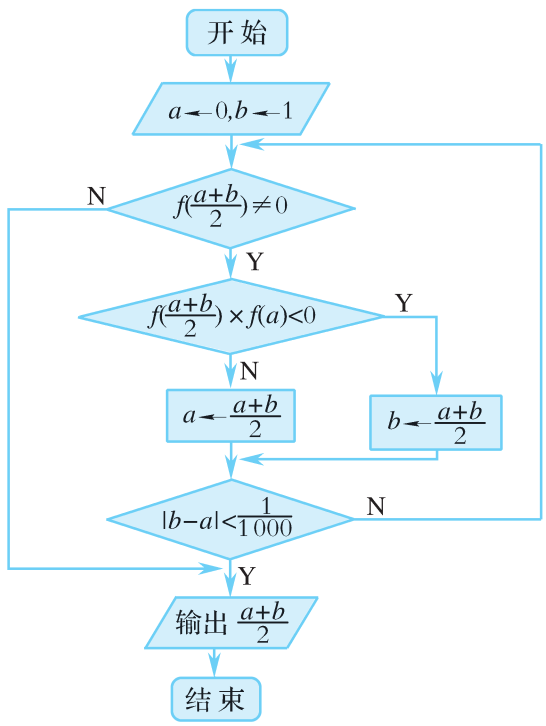


图 11－13

如此反复，直至区间的长度很小，使得区间内的任何一个值作为

根都能满足精度要求.

算法的伪代码如下：

S1　$a \leftarrow 0$，$b \leftarrow 1$

S2　While $f\left(\frac{a+b}{2}\right) \neq 0$ and $|b-a| \geqslant \frac{1}{1\ 000}$

S3　If $f\left(\frac{a+b}{2}\right) * f(a) < 0$ then $b \leftarrow \frac{a+b}{2}$

Else $a \leftarrow \frac{a+b}{2}$

S4　End while

S5　Print $\frac{a+b}{2}$

能用二分法计算函数曲线的交点吗？

你能想出比二分法更快的求根算法吗？

在社会生活和科技活动中，要处理很多数据，如学生的成绩，体育比赛的记录，城市的人口，图书目录，英汉字典中的单词等. 为了方便查找，需要把数据按某种规律排序.

这就提出了一些算法问题：

如何把散乱的一组数据按大小排列好（排序问题）？

如何把两组分别排好了的数据合并成一组有序的数据？

如何把一个新的数据插入到一组排好的数据之中（插入问题）？

插入问题，是有关排序的问题中最基本最简单的问题.

例 2　插入问题.

设 $a[1]$，$a[2]$，$a[3]$，…，$a[n]$，这 n 个数已经从小到大排序，现在要插入一个数 a，如何设计算法让计算机自动进行这个工作？

分析　类比思考找算法.

设想有 10 张在正面写了不同号码的纸牌，已经按号码大小在桌子上排成一行，但背面向上，你看不见号码，只知道右边的大，左边的小. 另外给你一张写有号码的纸牌，你如何把手中的纸牌按号码大小的顺序排进去？

如果不许翻开纸牌比较号码大小，当然不行.

问题在于，能不能少翻一些牌，更快地完成任务？

马上会想到：先翻最右边一张.

如果运气好，它比手中牌的号码小，就把手中的牌摆到最右边，任务完成.

如果运气没有那么好，它比手中牌的号码大，就把它向右移动一下，腾出一个空位，再翻开空位左边的牌，即由右而左的第二张牌.

如果运气好，第二张牌比手中牌的号码小，就把手中的牌摆到空位处，任务完成.

如果运气没有那么好，第二张牌也比手中牌的号码大，就把它向右移动一下，腾出一个空位，再翻开空位左边的牌，即由右而左的第三张牌.

这样一张一张由右到左翻开比较. 如果翻开的牌比手中牌的号码小，手中的牌就在当前的空位处就位，任务完成.

最不幸的情形，翻到最左边一张也比手中牌的号码大，就把这最后一张向右移动一下，把手中的牌放到最左边的位子，任务也完成.

请你总结提炼，写出算法、程序框图和伪代码. 在本章附录中有相应的 BASIC 程序供你参考.

课后思考和讨论

上面例题中那样挨个比较，运气不好的时候桌子上有多少张牌就要比较多少次，是不是笨了一点？

你能想出聪明一点的办法吗？

温故可以知新，类比可以发现新思路. 前面学过的求方程的根的二分法，能不能用来解决插入问题呢？

第一次就翻开自左而右的第 5 张牌，是不是更合理？

如果手中牌的号码比它大，它和左边的 4 张就不再考虑了，下一步再翻开自左而右的第 8 张牌. 这样，接下去至多再翻两张就行了.

如果手中牌的号码比它小，它和右边的 5 张就不再考虑了，下一步再翻开自左而右的第 2 张牌. 这样，接下去至多也是翻两张就行了.

想一想，如果桌子上有 100 张牌，用这种二分插入法至多只要翻开多少张牌？

例 3 排序问题.

要将数列 $a[1]$，$a[2]$，$a[3]$，…，$a[n]$，从小到大排序，如何设计算法让计算机自动做这个工作？

解法一 类比思考找算法.

设想有 10 张写了不同号码的纸牌，在桌子上排成一行，如何调换这些纸牌，使它们按号码排成自左而右一个比一个大的顺序？

容易想到，可以先取两张牌排好顺序，然后一张一张地插入，上面已经详细讨论了插入的算法，所以排序也就不成问题了！

如果 10 张牌的号码自左而右为

22，18，10，24，33，17，9，2，5，19

按插入法排序的过程和比较的次数就是：

[18,22],10,24,33,17,9,2,5,19 （前 2 张排序，比 1 次）

[10,18,22],24,33,17,9,2,5,19 （将 10 插入，比 1 次）

[10,18,22,24],33,17,9,2,5,19 （将 24 插入，比 2 次）

[10,18,22,24,33],17,9,2,5,19 （将 33 插入，比 2 次）

[10,17,18,22,24,33],9,2,5,19 （将 17 插入，比 3 次）

[9,10,17,18,22,24,33],2,5,19 （将 9 插入，比 3 次）

[2,9,10,17,18,22,24,33],5,19 （将 2 插入，比 3 次）

[2,5,9,10,17,18,22,24,33],19 （将 5 插入，比 3 次）

[2,5,9,10,17,18,19,22,24,33], （将 19 插入，比 3 次）

总共做了 21 次比较，完成排序.

解法二 另一种思路是：先找出一张冠军（最大的）放在最右边，下一步再在其余的 9 张牌中找出最大的放在右边第二位，再在其余 8 张中找出最大的放在右边第三位. 依次类推，直到剩下一张.

如何具体操作呢？

下面是一位计算机科学家提出的办法：

比较最左边的第一张和第二张，如果第二张号码大，则第一张原位不动，第二张为冠军候选人.

如果第一张号码大，则将两张牌调换位置，按新的位置，也是第二张为冠军候选人.

接下去，比较最左边的第二和第三张，如果第三张号码大，则第二张原位不动，第三张为冠军候选人.

如果第二张号码大，则将两张牌调换位置，按新的位置，也是第三张为冠军候选人.

接下去，比较最左边的第三张和第四张. 经过比较和调换，第四张成了冠军候选人.

这样冠军候选人就自左向右运动，直到浮出水面，成为真的冠军.

如果在上述过程中的每一步都没有调换牌的位置，说明所有的牌已经按左小右大排好了顺序. 任务完成.

若不然，再从左边的 9 张牌中，用同样的方法让亚军的候选人自左向右运动，直到浮出水面，成为真正的亚军.

如果在产生亚军过程中的每一步都没有调换牌的位置，说明所有的牌已经按左小右大排好了顺序. 任务完成.

然后，类似方法让第三名、第四名……依次出现，排在冠、亚军的后面. 如果在产生某一名次过程中的每一步都没有调换牌的位置，说明这时所有的牌已经按左小右大排好了顺序. 任务完成.

这些“候选人”自左向右运动并浮出的过程好像水中的气泡向上冒一样，所以这种算法就叫“冒泡法”.

仍以刚才的 10 个号码为例，冠军浮出水面的过程如下：

22,18,10,24,33,17,9,2,5,19	（初始数列）
(18,22),10,24,33,17,9,2,5,19	（交换 1，2 项）
18,(10,22),24,33,17,9,2,5,19	（交换 2，3 项）
18,10,(22,24),33,17,9,2,5,19	（3，4 项不必交换）
18,10,22,(24,33),17,9,2,5,19	（4，5 项不必交换）
18,10,22,24,(17,33),9,2,5,19	（交换 5，6 项）
18,10,22,24,17,(9,33),2,5,19	（交换 6，7 项）

18,10,22,24,17,9,(2,33),5,19　（交换 7，8 项）

18,10,22,24,17,9,2,(5,33),19　（交换 8，9 项）

18,10,22,24,17,9,2,5,(19,33)　（交换9,10项,选出冠军 33）

然后，用同样的办法从 9 个数中选出亚军.

当然，也可以让较小号码的牌从右向左运动，先选出最小的.

课后思考和讨论

把上面例子中的数列用冒泡法排下去，直到排好顺序，算一下，比较多少次?

多试几个例子，插入法和冒泡法哪一个更好?

想一想，有没有别的排序的方法呢?

可不可以把数列分成几段，分别排序，再合并起来呢?

例 4　韩信点兵（解一次同余方程组）.

汉代大将韩信，不但有杰出的军事才能，传说他还是一位数学天才.

一次，韩信在操练士兵. 士兵排成 3 列纵队时，最后只剩 1 人；排成 5 列纵队，剩 4 人；排成 7 列纵队，剩 3 人. 韩信略想一下，便责问身旁的值日副将，刚才你报告今天到场的士兵是 1 264 人，实际上怎么只有 1 249 人呢?

解　这是中国古代流传于民间的一道趣味算题. 解题的算法十分巧妙，人们称为“鬼谷算”，也叫“隔墙算”，或称为“韩信点兵”. 在国际数学界，至今以“中国剩余定理”之名广为人知.

到了明代，数学家程大位用诗歌概括了这一算法，他写道：

三人同行七十稀，
五树梅花廿一枝，
七子团圆月正半，
除百零五便得知.

歌诀里让人记着几个数：3 和 70，5 和 21，7 和 15，还有 105，即 $3\times5\times7$.

这些数的用法是：题中 3 列纵队剩 1 人，用 1 乘 70；5 列纵队剩 4 人，用 4 乘 21；7 列纵队剩 3 人，用 3 乘 15．三个乘积相加.

$$1\times70+4\times21+3\times15=199.$$

这个 199，是符合题中条件的士兵数目．因为 199 用 3 除余 1，5 除余 4，7 除余 3．但是，因为 105 是 3，5，7 的公倍数，所以 199 加上或减去若干个 105 仍然符合条件．于是，94，199，304，409，514，619，724，829，…都可能是答案．韩信根据现场观察和值日副将的报告，选择了和 1 264 最接近的解，即 $199+10\times105=1\ 249$.

歌诀里的 70，21，15 又是从何而来？原来：

70 是 5 和 7 的公倍数，并且除以 3 余 1；

21 是 3 和 7 的公倍数，并且除以 5 余 1；

15 是 3 和 5 的公倍数，并且除以 7 余 1.

而 3，5，7 的最小公倍数是 105.

5 和 7 的公倍数中，为什么恰巧能找到除以 3 余 1 的数呢？原因是 5×7 和 3 互素，即 5×7 和 3 没有大于 1 的公约数．只要两个整数 a 和 b 互素，a 的倍数中一定有被 b 除余 1 的数.

如果用含有字母的式子表示，就更清楚明白了．如果一个数分别除以 3，5，7 的余数是 a，b，c，则这个数可以表成：

$$70a+21b+15c-105n\ (n \text{ 为任意整数}).$$

题目中“排成 3 列纵队剩 1 人”，即除以 3 余 1，用数学式子表达，就是同余式：

$$x\equiv1(\bmod 3).$$

含有未知数 x 的几个同余式联立，构成同余式方程组，如

$$x\equiv2(\bmod 3),$$

$$x\equiv3(\bmod 5),$$

$$x\equiv4(\bmod 7).$$

问同时满足这三个同余方程的最小正整数 x 是多少？

答案是：$2\times70+3\times21+4\times15=263$，

$263-105\times2=53.$

所以这个最小正整数 x 是 53.

如果不知道用上面的歌诀构造的算法，可以用计算机硬算：从 4 开始，每次加 7，这样得到的数除以 7 自然余 4. 每加一个 7，就检查一下它是否满足除以 3 余 2 和除以 5 余 3 的条件，满足了就输出答案，不满足继续进行，这样也可以找到答案.

相信你能够写出程序框图和伪代码.

在“Z＋Z 超级画板”程序工作区运行下列程序可以得到答案：

```
m=4;
while((Mod(m,3)-2)^2+(Mod(m,5)-3)^2>0)
{m=m+7;}
```

注意，这段程序中字母大小写不能改变.

运行结果是 53.

你能画出对应这三行程序的框图吗？

课后思考和讨论

如果上面的问题中把 3，5，7 换成 5，7，11，歌诀中的 70，21，15 和 105 应当改成哪几个数呢？

从上面的几个案例看出，解决同一个问题，可以有不同的算法. 有的算法效率高，有的算法效率低.

找到算法是关键. 有了算法，写程序是技术问题. 只要细心、耐心、虚心地边学边做，总会成功.

找寻一个好的新算法，是数学和计算机科学中的创造性的劳动.

如果一时没有找到好的算法，写个简单的程序让计算机硬算，也可解燃眉之急.

习题 4

学而时习之

1. 编写一个程序框图，将下列数据从小到大排序. 给定的数据序列是：1，2，19，13，3.

2. 用基本算法语句写一个算法，输入 20 个数据后，输出其中正数、负数和零的个数.

3. 高二年级女篮队员的身高如下（单位：m）：

   1.68　1.62　1.72　1.74　1.65　1.71　1.75　1.71　1.67

   1.67　1.73　1.72　1.77　1.69　1.77　1.72　1.68　1.70

   设计一个算法，从这些身高数据中搜索出高于 1.70 m 者，并画出程序框图.

温故而知新

4. 给定一个有序的数表，设计一个查询重复数据出现次数的统计算法.

5. 给定两个有序的数表，设计一个将此二个数表合并成一个有序数表的算法.

   (1) 用文字叙述步骤；

   (2) 用框图表达算法；

   (3) 用基本算法语句写出伪代码.

数学实验

在计算机上体验编程(选学)

动手编写程序，在计算机上运行自己的程序，你对算法的理解就会更深刻．看到计算机执行你的计划，快速准确地给出问题的解答，你会有一种成就感．由于 BASIC 语言不能处理符号计算，对较大的数的运算也不能给出准确值，我们推荐你使用“Z＋Z 超级画板”（以下简称超级画板）的程序工作区来体验编程．打开超级画板，在左方工作区下部单击“程序”按钮，进入程序工作区．在这里可以实现我们学过的许多算法语句和例子．为了醒目起见，下面的程序用黑体．

1．赋值语句和定义函数

超级画板中的赋值语句和数学中常用的一样，用等号．要给 a 赋值 5，可键入

a＝5；

这里，分号表示一个语句的结束．

注意，输入程序要在英文输入状态下操作．

执行程序的操作方法，是把鼠标的光标放在分号后面，按着 Ctrl 键打 Enter 键（这是超级画板程序工作区中执行程序的操作方法，以下只说“执行”，不再解释），屏幕显示：≫5＃（计算机执行下面的语句时，从这个＃后开始阅读，所以这个＃是有用的）．这是计算机对所执行的程序的回答，叫作“返回”，表示已经将 a 赋值为 5．不信你再键入：

a＋3；

执行后返回

≫8；＃

这说明计算机已经知道 a 的当前值是 5．

如果要让 a 的值增加 2，可键入

a=a+2;

这行命令的含义是把 a 的当前值加 2 后作为 a 的新值，我们知道这也是**赋值语句**. 执行后返回：

≫7#

这表明 a 的当前值已经改变为 7. 如不放心，要确认，可键入

a;

执行后返回.

≫7#

现在将 b 赋值为 3，键入“b=3;”，执行，于是 a，b 都被赋值，a=7，b=3.

例 1 编写一段程序，使 a，b 交换所赋的值.

解 要有第 3 个变量作为过渡，才能实现交换. 程序为：

c=a;

a=b;

b=c;

执行上述程序，再检查一下，a 和 b 的当前值是不是已经交换了.

例 2 梯形上下底分别为 a=3，b=7，高 h=4；编写一段程序计算梯形面积 s.

解 程序为

a=3; b=7; h=4;

s=(a+b) * h/2;

注意，在程序语言中一般用 * 表示乘号，不能省略.

执行后返回，

≫20#

即梯形面积等于 20.

你会想，直接键入“s=(3+7) * 4/2;”，执行后不是一样吗，何必先给 a，b，h 赋值呢？

先给 a，b，h 赋值的好处在于，如果要计算其他梯形的面积，

只要复制这段程序，把前面的数据改一下就可以执行，而不必改动公式中的数据. 如果所用的公式比较复杂，这样先赋值再用公式计算的优越性就很明显了.

如果想再方便一些，可以把这段程序做成一个计算梯形面积的函数. 为此只要键入

s(a,b,h){(a+b) * h/2;}

执行后返回：

≫s (a, b, h) #

这说明，函数 s(a,b,h) 的定义已经完成. 这里，s 叫作**函数名**，a，b，h 叫作**变元或参数**，花括弧中的语句，可以是 1 行或几行，叫作**函数体**. 这是定义函数的一般方法.

要使用这个函数计算上下底分别为 a=3，b=7，高 h=4 的梯形面积，只要键入 s(3,7,4);

执行后就会返回答案，即 20.

例 3 编写一个由三角形三边为 a，b，c，计算其面积 m 的函数程序.

解 使用海伦公式，即秦九韶的三斜求积公式，可写成下列函数程序：

m(a,b,c){s=(a+b+c)/2;
(s * (s-a) * (s-b) * (s-c))^(1/2);}

执行后就建立了函数 m(a,b,c);要计算三边长为 5，6，7 的三角形面积，只要键入 m(5,6,7);执行即可.

例 4 编写解二元一次方程组的程序，并用来解下列方程组：

$$\begin{cases}3x+2y=5,\\4x-3y=18.\end{cases}$$

解 一般的二元一次方程组的形式是

$$\begin{cases}ax+by=e,\\cx+dy=f\end{cases}\quad(ab-bc\neq0).$$

用消元法解此方程组，得到

$$x=\frac{ed-bf}{ad-bc},$$

$$y=\frac{af-ce}{ad-bc}.$$

据此，可以分别写出计算 x，y 的函数程序：

x(a,b,c,d,e,f){(e*d-b*f)/(a*d-b*c);}

y(a,b,c,d,e,f){(a*f-c*e)/(a*d-b*c);}

执行后就建立了二元一次方程组求解函数．要解例中的具体问题，只要执行：

x(3，2，4，-3，5，18)；

就得到 3，即 x=3；若执行

y(3，2，4，-3，5，18)；

则得到-2，即 y=-2.

我们看到，赋值语句虽然简单，用它还是可以做不少事的.

上面所举的例子，都是把数字赋予字母变量．其实，也可以将字母或数学表达式赋予字母变量．如果键入

a=1+y；

执行后返回

≫y+1#

再键入

a^3；

执行后得到

≫y^3+3*y^2+3*y+1#

例 5 把(x+y)^9 的展开式看成 y 的多项式，写出求其中 y^4 项的系数的程序.

解 函数 Coeff(f,x,k)可以求出多项式 f 中 x 的 k 次项的系数，使用它容易写出所要程序：

p=(x+y)^9；

Coeff(p,y,4)；

执行后返回

```
≫126 * x^5 #
```

2. 条件语句

在超级画板提供的编程环境中，条件语句的一般格式是：

if （A） {B} else {C}

这里 **A** 是条件，**B** 是 **A** 成立时要执行的一些语句，**C** 是 **A** 不成立时要执行的一些语句．在花括弧内的 **B** 和 **C**，也可以是条件语句．

比起一般的伪代码，这里省略了(**A**)和{**B**}之间的"then"．

例 6 编写求一元二次方程

$$ax^2+bx+c=0 \quad (a\neq 0)$$

的实根的函数程序．

解 先判断是否有实根，有实根时再用公式求根．函数定义如下：

```
root(a,b,c,i)
{d=b^2-4*a*c;
if(d<0){No;}
else {if(i==1){(-b+d^(1/2))/(2*a);}
else {(-b-d^(1/2))/(2*a);}}}
```

若无实根返回"No"；有实根则当 i=1，2 时分别给出两根．

注意，判断条件 i=1 要写成 i==1，以区别于赋值语句．

执行后返回：

```
≫root(a, b, c, i) #
```

若方程为 $x^2-3x+2=0$，执行情形为：

```
root(1, -3, 2, 1);
≫2#
root(1, -3, 2, 2);
≫1#
```

无实根的情形如：

```
root(1, 1, 2, 1);
≫No#
```

利用条件语句，可以定义分段函数．

例 7　编写一个计算符号函数 sgn(x)的程序.

解　由 sgn(x)的定义可得：

sgn(x){if(x==0){0;}

else{if(x>0){1;}else{−1;}}}

请执行并检验是否正确.

将上列程序略加改动，可以定义其他分段函数.

例 8　设 M(a，b，c)是 a，b，c 三个实数中最大者，写出计算 M(a，b,c)的程序.

解　显然有

M(a,b,c){if(a>b){if(a>c){a;}else{c;}}

else{if(b>c){b;}else{c;}}}

也可以先定义一个函数从两个数中选出较大的，再用复合函数求出三个或四个数中最大的数来. 例如，一次执行下列四行程序：

M(a,b){if(a>b){a;}else{b;}}

M3(a,b,c){M(M(a,b),c);}

M4(a,b,c,d){M(M(a,b),M(c,d));}

M5(a,b,c,d,e){M(M(a,b),M3(c,d,e));}

就分别定义了从 2，3，4，5 个数中选出最大的数的函数. 若直接编写从 5 个数中选出最大数的函数 M5(a，b，c，d，e)的程序，要费事得多. 可见函数概念多么有用. 有了 M5，一下就能写出从 25 个数中选出最大数的程序!

例 9　设 m，n 是两个整数且 m>0，写出函数程序计算用 m 除 n 的余数. 也就是满足条件：

$$\mathbf{n=qm+r}\quad(\mathbf{q,r}\text{ 是整数},\mathbf{0\leqslant r<m})$$

的 r.

此函数一般记作 **Mod(m,n)**，叫作**模 n 求余函数**. 超级画板有此内置函数，为避免重名，我们将它取名为 mod1.

解　函数 floor(x)表示不大于 x 的最大整数. 利用它容易写出：

mod1(n,m){m * (n/m-floor(n/m));}

请执行并用算例检验.

例 10 闰年的 2 月是 29 天，比平年的 2 月多 1 天. 判断某一年是不是闰年，先看此年份能否被 4 整除. 不能整除就是平年；能被 4 整除但不能被 100 整除则为闰年；能被 100 整除时若又能被 400 整除则为闰年，否则是平年. 写出根据年份判断是否闰年的程序，并用来检验下列年份是否闰年：1816，1994，1800，2000.

解 调用函数 Mod(m,n)可以检验年份能否被 4，100，400 整除. 程序为：

```
Rn(m){if (Mod(m,4)>0) {No;}
      else {if (Mod(m,100)>0) {Yes;}
        else {if (Mod(m,400)==0) {Yes;}
          else{No;}}}}
```

执行后检验：运行 Rn(1816)；Rn(1994)；Rn(1800)；Rn(2000)；分别返回 Yes，No，No，Yes.

3. 循环语句

超级画板中的循环语句主要有 for 语句和 while 语句.

while 语句的格式为

while(条件){一些语句}

当条件满足时就顺次执行花括弧中的语句，每执行一轮就检查一次条件，当条件不满足时就退出循环，跳过花括弧执行后面的语句.

用 While 语句可以实现各种循环过程，不论是否预先知道循环次数都行.

例 11 用 while 语句编写求两个正整数 a，b 的最大公约数的程序.

解 用更相减损法写出求最大公约数的函数 gcd(a，b)，程序如下：

```
gcd(a,b){while((a-b)^2>0)
  {if(a>b){a=a-b;}else{b=b-a;}}}
```

执行后使用它，返回的是最后一次循环后的 a 或 b 的值：

```
gcd(56, 35);
≫7#
gcd(273, 147);
≫21#
gcd(7267, 6192);
≫43#
```

若用欧几里得辗转相除法，程序如下：

```
gcd(a, b)
{if(a<b){r=b;b=a;a=r;}
    while(a*b>0)
    {r=b;b=Mod(a,b);a=r;}}
```

你能说出这样编程的理由吗?

例 12 用 while 语句编写程序计算小于 n（n>1）的正奇数的平方和.

解 用 S(n)记所求的和，有

```
S(n){k=-1;
    s=0;
while(k+2<n){k=k+2;s=s+k^2;}}
```

执行后计算几个例子：

```
S(4);
≫10#
S(10);
≫165#
S(100);
≫166650#
S(1000);
≫166666500#
```

例 13 一个正整数被 3 除余 2，被 5 除余 3，被 7 除余 6. 编写程序求出此数的最小值.

解 从 6 开始，不断加 7，每次检查是否被 5 除余 3 且被 3 除

余 2，满足时即为所求. 程序如下：

```
n=6;
while((Mod(n,5)-3)^2+(Mod(n,3)-2)^2>0){n=n+7;}
```

如果要求被 3 除余 a，被 5 除余 b，被 7 除余 c，写成函数就是：

```
ty357(a,b,c){n=c;
while((Mod(n,5)-b)^2+(Mod(n,3)-a)^2>0)  {n=n+7;}}
```

此函数要求 a，b，c 是分别小于 3，5，7 的正整数. 执行后，运行下列命令得到：

```
ty357 (2, 3, 4);
≫53#
ty357 (2, 4, 3);
≫59#
ty357 (1, 2, 3);
≫52#
```

循环的次数可以事先确定时，可以用 for 语句. 超级画板编程环境中的 for 语句格式为：

```
for(i=a;i<b;i=i+d) {S}
```

其中圆括弧中的 i 是循环变量，也可以用其他字母表示循环变量. i=a表示 i 的值从 a 开始，a 为初值；i<b 说明了 i 的上限为 b；i=i+d表示 i 每次加 d，d 为步长. 花括弧中的 S 表示一些语句，叫作循环体. 程序对每个 i 的值顺次执行循环体中的语句.

例 14 用for 语句编写程序计算所有小于 100 的正奇数的倒数和.

解 循环变量的初值为 1，上限为 100，步长为 2. 程序为：

```
s=0;
for(k=1; k<100; k=k+2) {s=s+1/k;}
```

执行后返回为：

≫(2908397099990131201)/(990000000000000000)=2.93777#

计算所有小于 n 的正奇数的倒数和的函数程序为：

```
sq(n){s=0;for(k=1;k<n;k=k+2){s=s+1/k;}}
```

例 15 用 for 语句编写模拟抛掷均匀硬币并计算 n 次中出现正面的频率的函数程序.

函数 rand(0, 2) 产生 [0, 2] 上的一个随机数.

解 用 rand(0, 2)>1 表示一次抛掷出现正面，所要的程序为：

```
B(n){s=0;
    for(k=0;k<n;k=k+1){if(rand(0,2)>1){s=s+1;}}
      s/n;}
```

执行后再运行下列命令，可得模拟抛掷 10 000 次中出现正面的频率：

```
B(10000);
≫(5013)/(10000)=0.5013#
```

在算法案例中，我们给出了用二分法计算函数零点的程序框图和伪代码；下面给出超级画板中的程序.

例 16 编写用二分法求函数零点的程序，求函数 e^x+x 的零点.

解 令 $f(x)=e^x+x$，则 $f(-1)<0$，$f(0)>0$，在 -1 到 0 之间必有 $f(x)$ 的零点. 用二分法求此零点的程序并执行如下：

```
Float(1);
```

执行后返回：

```
    ≫计算结果显示浮点数#
    f(x){e^x+x;}
    a=-1; b=0;
    u=f(a); v=f(b);
    while(u*v<0&&b-a>0.000001)
    {c=(a+b)/2; w=f(c);
        if(w==0){u=w;}
        else{if(u*w<0) {v=w; b=c;}
                else{u=w; a=c;}}}
    (a+b)/2;
```

符号 “&&” 表示 “并且”.

执行后返回：

```
≫f(x)
    -(1134285926818847)/(2000000000000000)=-0.567143#
```

上述结果精确到小数点后 6 位.

小结与复习

一、内容提要

1. 算法的例子和含义：一元二次方程和二元一次方程组求解算法.

2. 算法程序框图的三种基本逻辑结构：顺序结构、条件分支结构、循环结构.

3. 算法程序语言的基本算法语句——输入语句、输出语句、赋值语句、条件语句、循环语句.

4. 重要算法案例：求最大公约数，二分法求根，多项式求值，插入和排序，中国剩余定理.

二、学习要求

1. 算法的例子和含义.

(1) 通过对解决具体问题过程与步骤的分析（如二元一次方程组求解等问题），体会算法的思想，了解算法的含义.

(2) 通过模仿、操作、探索，经历通过设计框图表达和解决问题的过程.

2. 程序框图.

在具体问题的解决中，理解程序框图的三种基本逻辑结构：顺序、条件分支、循环.

3. 基本算法语句.

经历将具体问题的程序框图转化为程序语句的过程，理解几种基本算法语句——输入语句、输出语句、赋值语句、条件语句、循环语句，进一步体会算法的基本思想.

4. 通过阅读中国古代数学中的算法案例，体会中国古代数学

对世界数学发展的贡献.

三、需要注意的问题

算法的基本思想、算法的基本结构、算法的基本语句是算法学习的重点，难点是循环结构.

1. 有关算法的概念必须通过实例来学习；通过模拟计算机游戏，具体理解算法执行的过程.

2. 注重对算法基本思想的理解.

算法是数学教科书中新出现的内容，其思想是非常重要的，但是并不神秘，对于大家也不陌生. 解方程的算法、解不等式的算法、因式分解的算法……都是我们熟知的内容.

通过本模块的学习，要建立数学中的算法和计算机技术之间的联系，形式化地表示算法，有条件的最好能在计算机上实现. 为了有条理地、清晰地表达算法，往往需要将解决问题的过程整理成程序框图；为了能在计算机上实现，还需要将自然语言或程序框图翻译成计算机语言.

3. 算法不仅是本章的学习内容，其思想方法也贯穿在所有的数学课程之中. 在数学学习过程中，提倡大家有意识地运用算法知识解决相关问题，使程序化思想成为思考问题的习惯.

复习题十一

学而时习之

1. 先填写解题步骤表，再编写解下列三元一次方程组的算法的伪代码.

$$\begin{cases} x+2y+3z=2, \\ 2x+y+z=1, \\ 3x-y+3z=4. \end{cases}$$

步　骤	目　　的	运　算
第一步		
第二步		
第三步		
第四步		

2. 7，8，9 三个月的平均气温是 24 ℃，7 月的平均气温是 8，9 两个月的平均气温之差的 4 倍，而 8 月份的平均气温比 7，8 两个月的平均气温高 4.5 ℃. 问 7，8，9 三个月的平均气温各是多少？
3. 求 2 759，2 573 的最大公约数.
4. 给定一条边（边长＞5），以此边长为斜边，作一个直角三角形，使其一个直角边等于 5. 写出作图的算法.
5. 求 45，75 的最小公倍数.

温故而知新

6. 甲乙两人玩棋子，桌上共有 30 个棋子，每人一次限取一子或二子，谁把最后一子取完后谁就是胜家. 如果甲先取子，请你判断谁必胜.（用含循环步骤的解题方案写出，要求条理明晰.）
7. 一队士兵 100 人同在小河的左岸，小河岸边有一只小船，船上坐着两个小朋友. 虽然每一个士兵和小朋友都会划船，但是小船容量有限，每次只能载一个大人或两个小孩. 这两个小朋友能否帮助这一队士兵过河？请你设计一个算法（方案）让他们实施.
8. 设计算法，求 $x^3+5=0$ 的一个近似根（精确到小数点后第二位）.
9. 随意写 100 个数字，其中包含两位连写是 53 者有多少（也许没有）？编写一个查询程序.

上下而求索

10. 在信息技术课老师指导下，写出排序问题的一个算法的伪代码，有条件时用 BASIC 语言或其他语言在计算机上实现.
11. 用“Z＋Z 超级画板”编写二分法求两个函数曲线交点并画图的程序，在计算机上实现.

附　与本章内容有关的一些 BASIC 程序

下面是本章中的伪代码和一些例子对应的 BASIC 程序，所有程序都在 QB4.5 环境下运行过.

将程序和伪代码对照，有助于理解有关的算法.

(1) 11.3.1 节正文（第 15 页）

用更相减损法求 a，b 的最大公约数之一：

```
10  INPUT a, b
20  IF a=b THEN 60 ELSE 30
30  IF a>b THEN 40 ELSE 50
40  a=a-b
    GOTO 20
50  b=b-a
    GOTO 20
60  PRINT a
```

(2) 11.3.1 节课后思考和讨论（第 16 页）

用更相减损法求 a，b 的最大公约数之二：

```
10  INPUT a, b
20  IF a=b THEN 50
30  IF a>b THEN a=a-b ELSE b=b-a
40  GOTO 20
50  PRINT a
```

(3) 11.3.2 节例 1（第 18 页）

计算梯形面积：

```
INPUT a, b, h
PRINT (a+b)*h/2
```

(4) 11.3.2 节例 2（第 18 页）

多项式求值：

算法一：

```
10  x=11
20  PRINT 5*x^3-3*x^2+8*x-13
```

算法二：

```
10  x=11
20  PRINT ((5*x-3)*x+8)*x-13
```

算法二的另一种写法：

```
10  x=11
20  f=((5*x-3)*x+8)*x-13
30  PRINT f
```

（5）11.3.3 节例 1（第 20 页）

判断长为 a，b，c 的 3 条线段能否构成三角形

```
10  INPUT a，b，c
20  IF a+b <=c THEN 60
30  IF a+c<=b THEN 60
40  IF b+c<=a THEN 60
50  PRINT "can"
    END
60  PRINT "can_not"
```

（6）11.3.3 节例 2（第 20 页）

计算水费：

```
10  INPUT t
20  IF t<=3 THEN
        s=1.1*t
    ELSE IF t<=5 THEN
        s=3.3+1.6*(t-3)
    ELSE s=3.3+2.2*(t-3)
    END IF
30  PRINT s
```

（7）11.3.4 节例 1（第 22 页）

用 while 循环语句实现更相减损法求最大公约数：

```
10  INPUT a, b
20  WHILE a<>b
30  IF a>b THEN a=a-b ELSE b=b-a
40  WEND
50  PRINT a
```

(8) 11.3.4 节例 2（第 23 页）

计算小于 1 000 的正奇数的立方和：

```
10  s=0
20  FOR k=1 TO 999 STEP 2
30  s=s+k^3
40  NEXT k
50  PRINT s
```

(9) 11.3.4 节后面“多知道一点”中的例（第 24 页）

用 until 循环语句实现更相减损法求最大公约数：

```
10  INPUT a, b
20  DO
30  IF a>b THEN a=a-b ELSE b=b-a
40  LOOP UNTIL a=b
50  PRINT a
```

用 while 循环语句实现辗转相除法求最大公约数：

```
10  INPUT a, b
20  WHILE a*b*(a-b) <>0
30  IF a>b THEN
        a=a MOD b
    ELSE b=b MOD a
    END IF
40  WEND
50  IF a=0 THEN a=b
60  PRINT a
```

(10) 11.4 例 1 (第 26 页)

用二分法求方程 x^5+x-1=0 根的近似值:

```
10   a=0
20   b=1
30   c=(a+b)/2
40   fc=c^5+c-1
50   WHILE fc<>0 AND b-a>1 /1000
60   IF fc>0 THEN
        b=(a+b)/2
     ELSE a=(a+b)/2
     END IF
70   c=(a+b)/2
80   fc=c^5+c-1
90   WEND
100  PRINT c
```

(11) 11.4 例 2 (第 27 页)

将数 a 按大小顺序插入到有序数列 a(i) 中的基本算法:

```
10   INPUT n, a
20   FOR i=1 TO n
30   INPUT a(i)
40   NEXT i
50   k=n
60   WHILE k>0 AND a<=a(k)
70   a(k+1)=a(k)
80   k=k-1
90   WEND
100  a(k+1)=a
110  FOR i=1 TO n+1
120  PRINT a(i)
130  NEXT i
```

将数 a 按大小顺序插入到有序数列 a(i) 中的二分快速算法：

```
10  INPUT n, a
15  FOR i=1 TO n
20  INPUT a(i)
25  NEXT i
26  IF a<=a(1) THEN k=1
27  IF a<=a(1) THEN 80
28  IF a >=a(n) THEN a(n+1)=a
29  IF a >=a(n) THEN 85
30  u=1
35  v=n
40  d=(u+v) MOD 2
45  m=(u+v-d)/2
50  WHILE v-u>1 AND a<>a(m)
55  IF a>a(m) THEN u=m ELSE v=m
60  d=(u+v) MOD 2
65  m=(u+v-d)/2
70  WEND
75  IF a=a(m) THEN k=m ELSE k=v
80  FOR i=n TO k STEP -1
81  a(i+1)=a(i)
82  NEXT i
83  a(k)=a
85  FOR i=1 TO n+1
90  PRINT a(i)
95  NEXT i
```

说明：执行程序时，先键入数列的总项数 n 和数 a，然后逐个键入 a(1)，a(2)，…，最后计算机顺序输出插入数 a 后的数列.

注意，有些 BASIC 程序环境限制数列的总项数不超过 10 项，即 $n<10$.

（12）11.4 节例 3（第 29 页）

用二分快速插入法排序：

```
10  INPUT p
11  FOR i=1 TO p
12  INPUT a(i)
13  NEXT i
14  IF a(1)>a(2) THEN
      c=a(1)
      a(1)=a(2)
      a(2)=c
    END IF
15  FOR n=2 TO p-1
16  a=a(n+1)
26  IF a<=a(1) THEN k=1
27  IF a<=a(1) THEN 80
29  IF a >=a(n) THEN 85
30  u=1
35  v=n
40  d=(u+v) MOD 2
45  m=(u+v-d)/2
50  WHILE v-u>1 AND a<>a (m)
55  IF a>a (m) THEN u=m ELSE v=m
60  d=(u+v) MOD 2
65  m=(u+v-d)/2
70  WEND
75  IF a=a(m) THEN k=m ELSE k=v
80  FOR i=n TO k STEP -1
81  a(i+1)=a(i)
82  NEXT i
83  a(k)=a
```

```
85  NEXT n
87  FOR i=1 TO p
90  PRINT a(i)
95  NEXT i
```

用冒泡法排序：

```
10  INPUT n
20  FOR i=1 TO n
30  INPUT a(i)
40  NEXT i
50  k=n-1
60  WHILE k>0
65  c=0
70  FOR j=1 TO k
76  IF a(j)>a(j+1) THEN
      b=a(j)
      a(j)=a(j+1)
      a(j+1)=b
      c=c+1
    END IF
77  NEXT j
79  IF c=0 THEN k=0 ELSE k=k-1
80  WEND
85  FOR i=1 TO n
90  PRINT a(i)
95  NEXT i
```

（13）11.4 节例 4（第 31 页）

求同余式方程组的最小解：

下列程序要求输入分别小于 3，5，7 的正整数 a，b，c，输出满足条件“被 3 除余 a，被 5 除余 b，被 7 除余 c”的最小正整数.

```
10  INPUT a，b，c
```

```
20  n=c
30  x=n MOD 3
40  y=n MOD 5
50  WHILE (x-a)^2 +(y-b)^2>0
60  n=n+7
70  x=n MOD 3
80  y=n MOD 5
90  WEND
95  PRINT n
```

数学文化

中国古代数学的算法思想和典型案例

中国古代数学的指导思想是以算法为纲，寓理于算．为解决一个数学问题，一般总是为其设计编制一个巧妙有效的算法．这种算法或是由加、减、乘、除与开方组成的有限次运算的公式，或是明确何时应做何种操作的一种所谓行为算法．各种算法都具有构造性的特点，与现代计算机上的证明与计算的编程不谋而合．

中国古代数学的算法思想在《九章算术》（刘徽注）、《孙子算经》、《海岛算经》（刘徽著）、《张邱建算经》、《数书九章》（秦九韶著）、《详解九章算术》（杨辉著）、《测圆海镜》（李冶著）、《四元玉鉴》（朱世杰著）以及祖家父子割圆求 π 和杨辉构作幻方等数学经典和重大数学成果当中，体现得淋漓尽致．

《九章算术》中 246 题几乎每题后面都有“术曰”一项，给出该题的具体算法．

例如 76 题：“今有出钱五百七十六，买竹七十八个，欲其大小率之，问各几何？”

“术曰”中指出①$576\div78$ 商 7 余 30；②挑出较大的 30 根竹子，每根付 $7+1=8$ 钱；③$78-30=48$，这 48 根较小的竹子每根付款 7 钱．

如果按现在通行的解法，设大竹 x_1 根，每根 y_1 钱，小竹 x_2 根，每根 y_2 钱，则得四元二次不定方程组（解不唯一）

$$\begin{cases}x_1+x_2=78,\\x_1y_1+x_2y_2=576.\end{cases}$$

解这个方程不是一件轻松的事．相比之下，我国古代的上述算法是何等的简洁和机智，真可以称为“绝招”，令人叹为观止．

再例如185题：“今有共买物，人出八，盈三；人出七，不足四．问人数物价各几何?”

若记人数为 x，物价为 y，《九章算术》中给出的算法是

$$x=\frac{b_1+b_2}{a_1-a_2},\qquad y=\frac{a_1b_2+a_2b_1}{a_1-a_2}.$$

其中出钱 a_1 盈 b_1，出钱 a_2 不足 b_2．这种算法史称“中国算法”．

《九章算术》中还有所谓“遍乘直除”的线性（一次）代数方程组的消元算法和勾股数等的重要算法．

《孙子算经》（成书公元元年左右，作者已不可考）上有名题“物不知数”（俗称“韩信点兵”）：

“今有物不知其数，三三数之余二，五五数之余三，七七数之余二，原物几何?”

宋代数学家把此题之算法写成了口诀诗（一），明代数学家程大位又把它改写成妇孺能通的算法口诀诗（二）：

（一）	（二）
三岁孩儿七十稀，	三人同行七十稀，
五留廿一事尤奇，	五树梅花廿一枝，
七变上元重相会，	七子团圆正月半，
寒食清明便得知．	除百零五便得知．

具体运算过程是：①$2\times70+3\times21+2\times15=233$，②$233-105=128$，③$128-105=23$．23为所求的最小数．一般而言，总结出“孙子定理”和秦九韶的“大衍求一术”，又称“中国剩余定理”，是数学史上的著名算法之一．

我国数学家不仅长于计算型的算法，而且也长于“行为算法”．例如杨辉推出的奇阶幻方的构作算法，所谓幻方是用1到 n^2 这 n^2 个数组成的 $n\times n$ 的数字方阵，每行、每列和两条对角线上的数字分别求和，所得之和一致．杨辉解读了《易经》（公元前2 000多年）上的“洛书”，得出高阶幻方的构造算法：

杨辉的算法是①画 $(2k+1)\times(2k+1)$ 个小方格的正方形；②在正方形的每边向外作一个由小方格垒成的等腰直角三角形；③把

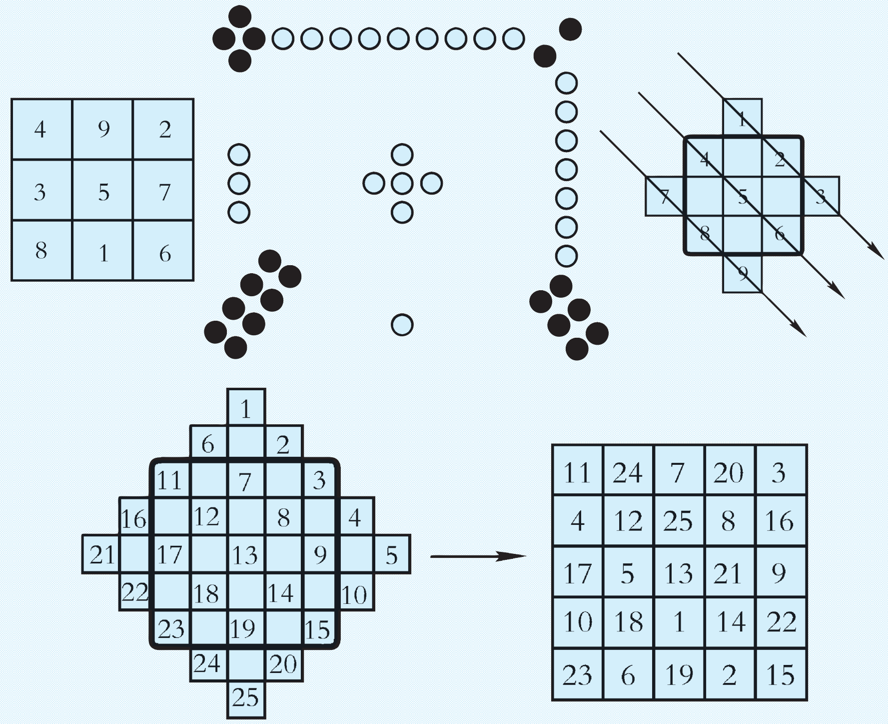


图 11－14

1，2，3，…，$(2k+1)\times(2k+1)$从左向右下方 45°走向依次抄入②中的一些方格，使之形成一个正方形数字阵；④把①中正方形之外的数收入内部，上方的下降 $2k+1$ 格，下方的上升 $2k+1$ 格，右方的左移、左方的右移 $2k+1$ 格.

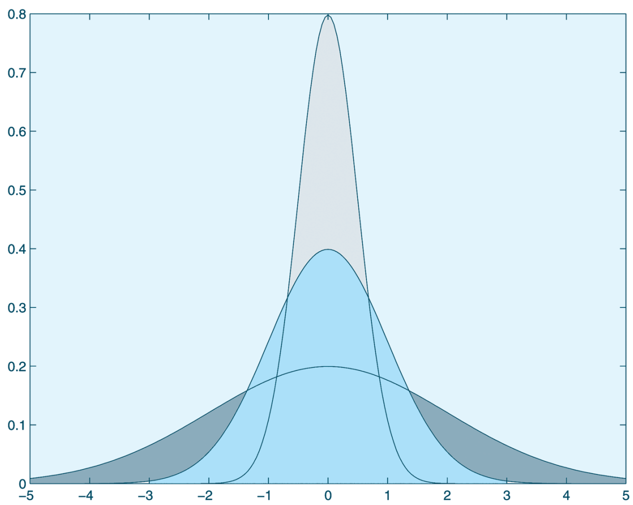
杨辉的上述构作幻方的算法真乃鬼斧神工.

我国古代数学中的算法不但追求巧妙漂亮，而且追求其有效性，即在合理的（不太长）的时间内能把算法执行完. 这种思想与今日的计算时间复杂度理论中的有效性思想一脉相承.

我国现代数学家吴文俊等继承了中国古代的算法思想且发扬光大之，成功地进行了几何定理机器证明的理论建设、算法设计和实际应用. 中国古代数学中的算法思想是中华民族伟大的数学成就之一，它具有宝贵的理论价值和实际意义.

第 12 章 统计学初步

数据纷繁沙一盘，　管窥蠡测理当然．
总体抽样图良策，　均值方差求指南．
辨明真假成功路，　分清主次艳阳天．
国运民生关统计，　知风知浪好行船．



高　斯（Gauss）研究测量误差时发现了正态分布曲线，这是对统计学的重要贡献．

现代生活是建立在数据之上的．没有数据，一切很难想象．统计学就是研究如何从数据中提取有用信息的科学，内容包括如何收集和分析数据．基于统计学的数据处理方法称为统计方法．

统计学初步仅仅帮助你了解统计学的一些基本语言，知道一些统计学的基本概念．学习了统计学初步后你也许会觉得知道一些统计学的初步知识是有用的．我们认为统计学还可以激发你的智力，给你的生活带来更多的乐趣．

12.1 总体和个体

日常生活中我们总是自觉或不自觉地和总体与样本打交道．夏天买西瓜时，先要看看这批西瓜甜不甜．如果瓜甜又不很贵，你可能买一个或两个．

我们可以称这批西瓜是一个总体，单个的西瓜是个体，但是这样就不能强调我们关心的是西瓜的甜度．因为西瓜的好坏还有其他的指标，例如个的大小，是否新上市的等等．

在关心这批西瓜的甜度时，我们称单个西瓜的甜度是“个体”，称所有的西瓜的甜度为“总体”．这样就把西瓜的甜不甜数量化了．

要了解一批西瓜的甜度情况，你不可能品尝每个西瓜．你只能买一两个吃一吃．然后通过这一两个西瓜的甜度判断这批西瓜的甜度．这就是用少数个体推断总体．我们把买的西瓜的甜度称为“样本”，于是你已经可以用样本推断总体．

12.1.1 总体、个体和总体均值

要调查全校期中考试的数学成绩时，称全校同学的期中数学成绩是总体，称单个同学的数学成绩是个体．

总体、个体和均值是统计学的最基本概念．

要调查全校同学期中考试的语文成绩时，称全校同学的语文成绩是总体，称单个同学的语文成绩是个体．

在统计学中，我们把所要调查对象的全体叫作**总体**（population），把总体中的每个成员叫作**个体**（individual）．

总体中个体的某一特征总可以用数量表示．为了叙述的简单和明确，我们把个体看成数量，把总体看成数量的集体．

调查全校同学期中成绩时，指出数学或语文是为了明确总体．不同的总体不能混为一谈．

总体中个体的数目有时是确定的，有时较难确定．调查全校同学期中考试的数学成绩时，参加考试的人数是明确的，相应总体的个数也就明确了．在调查全国人口的年龄分布时，总体是全国人口的年龄，是明确的，但是个体总数很难精确下来．

全校同学期中数学考试成绩的平均值是总体平均，全校同学期中语文考试成绩的平均值也是总体平均．总体平均是总体的指标，是我们所关心的指标．

总体平均是总体的平均值，也称为总体**均值**（mean）．

在统计学中，常用 μ（音 miu）表示总体均值．当总体含有 N 个个体，第 i 个个体是 y_i 时，总体均值

$$\mu=\frac{y_1+y_2+\cdots+y_N}{N}.$$

在判断一批西瓜甜不甜时，你没有必要知道一共有多少个西瓜．

练习

用 $\bar{x}$ 表示数据 x_1，x_2，…，x_n 的均值，用 b 表示常数．对于数据

$$y_1=x_1+b,\quad y_2=x_2+b,\quad \cdots,\quad y_n=x_n+b$$

的均值 $\bar{y}$，验证：

$$\bar{y}=\bar{x}+b.$$

练习的结论表明：每个数据增加相同的量，数据的均值也增加相同的量．

习题 1

学而时习之

1．简述总体平均的含义．

2．用 $\bar{x}$ 表示观测数据 x_1，x_2，…，x_n 的均值，用 a 表示常数．用 $\bar{y}$ 表示观测数据 $y_1=ax_1$，$y_2=ax_2$，…，$y_n=ax_n$ 的均值时，证明：

$$\bar{y}=a\bar{x}.$$

习题 2 的结论表明：数据同时扩大 a 倍，均值也扩大 a 倍．

3．对于数据

65	60	59	60	53	60	62	63
70	59	65	66	67	56	63	63
55	57	68	61	56	66	65	57

用 μ_1，μ_2，μ_3 分别表示第 1，第 2，第 3 行的平均值，用 μ 表示全体 24 个数的平均值.

（1）计算 μ_1，μ_2，μ_3 和 μ；

（2）是否有 $\mu=(\mu_1+\mu_2+\mu_3)/3$？

12.1.2　样本与样本均值

要了解一盘菜炒得好吃不好吃，你品尝一下就可以下结论了．没有必要等到把菜吃完再作结论．你品尝的菜就是样本，你的品尝就是把样本进行平均，然后你用样本的平均推断总体的平均.

考察 A 中学高一年级 500 个同学某时间的平均身高 μ．要得到这 500 个同学的平均身高不是一件很困难的事情，只要了解了每个同学的身高就可以利用公式

$$\mu=\frac{\text{这 500 个同学身高之和}}{500}$$

计算得到.

同一天对每个同学进行一次身高测量可以得到均值 μ 的准确值，但是要花费老师和同学们较多的时间和精力．统计上解决这类问题的最好方法是进行抽样调查，例如在 500 个同学中只具体测量 50 个同学的身高，用这 50 个同学的平均身高作为总体平均身高 μ 的近似．这时我们称这 50 个同学的身高为总体的一个样本，称 50 为样本量.

从总体中抽取一部分个体，称这些个体为**样本**（sample）.

样本也叫作**观测数据**（observed data）.

称构成样本的个体数目为**样本容量**，简称为**样本量**（sample size）.

称从总体抽取样本的工作为**抽样**（sampling）.

按照上面的定义，总体也是一个样本，称为**全样本**，但是样本一般不是总体.

在考虑身高问题时，对于前述被选中的 50 个同学，用 x_1，x_2，…，x_{50} 分别表示第 1，第 2，…，第 50 个同学在调查日的身高，则这 50 个同学的身高

$$x_1,\ x_2,\ \cdots,\ x_{50}$$

是样本. 用 n 表示样本量，则 $n=50$.

样本均值是样本的平均值，用 $\bar{x}$ 表示.

如果样本是 y_1，y_2，…，y_n，就用 $\bar{y}$ 表示样本均值.

总体均值是总体的指标，是一个固定的量. 但是样本均值依赖于样本的选择，从不同的样本会计算出不同的样本均值. 所以我们说样本均值带有随机性.

和总体均值 μ 相比较后知道，只要抽样合理，对于较大的样本量 n，样本均值 $\bar{x}$ 会接近 μ. 于是，$\bar{x}$ 是总体均值 μ 的近似，所以称为 μ 的**估计**（estimator）.

问题 在考察 A 中学高一年级 500 个同学的平均身高时，决定调查 50 个同学，用这 50 个同学的平均身高作为全年级平均身高的估计. 有 5 位女同学主动承担了这次调查任务，她们每人负责选择了 10 个同学，在 9 月的第一周测量出了所选择的 50 个同学的身高如下（单位：cm）：

156	166	165	157	160	164	162	158	158	164
155	165	165	172	165	158	164	155	161	161
162	160	168	150	164	167	166	165	162	166
165	160	159	160	153	160	162	163	170	159
165	166	167	156	163	163	155	157	168	161

其中 $x_1=156$，$x_2=166$，$x_3=165$，…，$x_{50}=161$ 分别是第 1，第 2，…，第 50 个被选中的同学的身高，样本量 $n=50$. 对上述第 50 个观测数据进行平均后得到

$$\bar{x}=\frac{x_1+x_2+\cdots+x_n}{n}=161.86\ (\text{cm}).$$

于是，对全年级平均身高 μ 的估计是 161.86 cm.

上述调查结果公布后，引起了同学们的议论，普遍认为

161.86 cm偏小了．问题出在哪里呢？我们在学习抽样调查方法时再解答这个问题．

练　习

用 $\bar{x}$ 表示观测数据 x_1，x_2，…，x_n 的均值，用 a，b 表示常数．用 $\bar{y}$ 表示观测数据 $y_1=ax_1+b$，$y_2=ax_2+b$，…，$y_n=ax_n+b$ 的均值时，证明 $\bar{y}=a\bar{x}+b$.

习题 2

学而时习之

1. 简述样本均值和总体均值的关系．
2. 当样本 x_1，x_2，…，x_n 中有 n_1 个 y_1，n_2 个 y_2，…，n_k 个 y_k，验证：
$$\bar{x}=\frac{n_1y_1+n_2y_2+\cdots+n_ky_k}{n}.$$
3. 将某调查公司得到的 20 个数据从小到大排列后得到

680	680	680	680	685	685	685	685	685	690
690	690	690	690	690	696	696	696	700	700

计算样本均值 $\bar{x}$.

4. 将一个总体中的 $5n$ 个个体平均分成 n 份，每份 5 个个体．先计算每份的均值，得到 n 个均值．这 n 个均值的平均值是否等于总体均值？证明你的结论．

12.1.3　方差和标准差

拔河比赛是一项有益于身体健康和增进团结的体育活动，某居民区的 2 号楼和 6 号楼决定进行拔河比赛．2 号楼组成 2 号队，6 号楼

组成 6 号队，每队 15 人．参加比赛时，2 号队的年龄组成是

8，8，9，9，9，10，10，10，

10，12，57，61，62，65，65．

6 号队的年龄组成是

26，26，26，26，26，27，27，27，

27，27，28，28，28，28，28．

这两个队的平均年龄都是 27 岁，但是各队一出场，大家就基本能够判断出比赛的结果了．2 号队的年龄相差悬殊，是老爷爷带小朋友；6 号队的年龄整齐，都是中青年．看来只靠平均年龄无法判定拔河队的实力，还需要有一个能衡量年龄的整齐程度的量．这个量就是要学习的方差．

1．总体方差．

当 y_1，y_2，…，y_N 是总体的全部个体，μ 是总体均值时，称

$$\sigma^2=\frac{(y_1-\mu)^2+(y_2-\mu)^2+\cdots+(y_N-\mu)^2}{N}$$

是总体的平均平方误差，简称为**总体方差**或**方差**（variance）．

总体方差 σ^2 描述了总体中的个体向总体均值 μ 的集中程度：方差越小，个体向 μ 集中得越好．

总体方差 σ^2 也描述了总体中个体的整齐程度或波动幅度，方差越小，个体就越整齐．

σ 音 sigma，σ^2 读作 sigma 方．

可以计算 2 号拔河队年龄的方差是 $\sigma_2{}^2\approx 616.3$ 号拔河队年龄的方差 $\sigma_6{}^2\approx 0.7$．相差太悬殊了．

例 1 同一年级的甲班有 35 个同学，乙班有 37 个同学．期中考试后，数学的平均成绩分别是 79.4 分和 82.7 分．方差分别是 68.6 和 148.8．如何就这次考试的结果评价这两个班的数学课的学习情况．

解 从平均分上看，乙班的数学平均成绩好于甲班，但是从学习的整齐程度方面看，甲班好于乙班．甲班的分数比乙班的分数更集中．

下面是这两个班数学考试成绩的从小到大排列．

甲班成绩

60	65	65	66	70	71	71	72	72	75	76	78
79	79	80	80	81	81	81	82	82	83	83	83
84	84	85	85	85	85	86	87	91	95	98	

乙班成绩

52	56	64	67	67	67	69	73	74	74	77	77	80
82	83	83	83	83	85	85	86	87	88	88	88	88
89	90	90	92	97	98	99	99	100	100	100		

从中看出，甲班没有同学不及格，也没有同学得满分. 乙班有同学得满分，但是也有同学不及格. 甲班的数学成绩更整齐.

2. 样本方差.

知道总体方差后，可以作出更好的比较结果.

给定 n 个观测数据 x_1，x_2，…，x_n，用 $\bar{x}$ 表示这 n 个数据的均值. 称

$$s^2=\frac{1}{n}[(x_1-\bar{x})^2+(x_2-\bar{x})^2+\cdots+(x_n-\bar{x})^2]$$

为这 n 个数据的**样本方差**，也简称为**方差**.

样本方差 s^2 是描述观测数据关于样本均值 $\bar{x}$ 发散程度的指标，也是描述数据的发散程度或波动幅度的指标.

样本方差依赖于样本的选取，也带有随机性. 样本方差是总体方差的估计.

例 2 一箱内有 50 个苹果，净重 10 kg，则苹果的平均重量是

$$\frac{10\ 000}{50}=200\ (\mathrm{g}).$$

要了解这箱苹果的整齐程度，就需要估计这箱苹果重量的方差 σ^2.

解 从中抽出 10 个，测得这 10 个苹果的重量（单位：g）是

201，218，187，192，193，198，202，194，176，291.

样本均值是

$$\bar{x}=\frac{201+218+\cdots+291}{10}=205.2\ (\mathrm{g}).$$

样本方差是

$$s^2=\frac{1}{10}[(201-205.2)^2+(218-205.2)^2+\cdots+(291-205.2)^2]$$
$$=923.76(\mathrm{g}^2).$$

我们可以用样本方差 $s^2=923.76$ （g^2） 作为总体方差 σ^2 的估计.

3. 方差的计算公式.

$$s^2=\frac{1}{n}(x_1{}^2+x_2{}^2+\cdots+x_n{}^2)-\bar{x}^2.$$

证 将每个$(x_i-\bar{x})^2$展开，再利用

$$x_1+x_2+\cdots+x_n=n\bar{x},$$

得到

$$\begin{aligned}s^2&=\frac{1}{n}[(x_1-\bar{x})^2+(x_2-\bar{x})^2+\cdots+(x_n-\bar{x})^2]\\&=\frac{1}{n}[(x_1{}^2+x_2{}^2+\cdots+x_n{}^2)-2(x_1+x_2+\cdots+x_n)\bar{x}+n\bar{x}^2]\\&=\frac{1}{n}[(x_1{}^2+x_2{}^2+\cdots+x_n{}^2)-2n\bar{x}\bar{x}+n\bar{x}^2]\\&=\frac{1}{n}[(x_1{}^2+x_2{}^2+\cdots+x_n{}^2)-n\bar{x}^2]\\&=\frac{1}{n}(x_1{}^2+x_2{}^2+\cdots+x_n{}^2)-\bar{x}^2.\end{aligned}$$

4. 标准差.

在例 2 中，数据的单位是 g，样本方差 s^2的单位是 g^2，和数据的单位不一致. 为了使描述数据的波动幅度的量和数据的单位一致，我们再引入标准差.

标准差（standard deviation）是方差的算术平方根；

如果 s^2是样本方差，就称 $s=\sqrt{s^2}$是**样本标准差**；

如果 σ^2是总体方差，就称 $\sigma=\sqrt{\sigma^2}$是**总体标准差**.

在例 2 中，10 个苹果重量的标准差是 $s=\sqrt{923.76}=30.39$（g）. 其单位和数据的单位一致.

当数据带有单位时，标准差的单位是和数据的单位一致的. 标准差也是描述数据发散程度或波动幅度的指标. 样本标准差是总体标准差的估计.

给定数据 x_1，x_2，…，x_n和均值 $\bar{x}$. 由方差计算公式知道，标准差 s 可以由下面的公式之一计算，

$$s=\sqrt{\frac{1}{n}[(x_1-\bar{x})^2+(x_2-\bar{x})^2+\cdots+(x_n-\bar{x})^2]},$$

$$s=\sqrt{\frac{1}{n}(x_1{}^2+x_2{}^2+\cdots+x_n{}^2)-\overline{x}^2}.$$

例 3 比赛中，甲乙两位射击运动员分别进行了 10 次射击，成绩分别如下：

甲：9.5　9.9　9.9　9.9　9.8　9.7　9.5　9.3　9.6　9.6

乙：9.4　9.3　9.5　9.0　9.1　9.8　9.7　9.5　9.3　9.4

问哪个运动员平均水平高，哪个运动员水平更稳定.

解 用 $\overline{x}$，s_x 和 $\overline{y}$，s_y 分别表示甲和乙成绩的均值和标准差. 经过计算得到

$$\overline{x}=9.67,\quad s_x=0.195\ 2,\quad \overline{y}=9.4,\quad s_y=0.232\ 4.$$

甲的平均水平和稳定性都比乙好.

练习

练习的结论表明：观测数据同时加上相同的常数后发散程度不变，方差也不变.

用 s_x^2 表示 x_1，x_2，…，x_n 的方差，用 b 表示常数，用 s_y^2 表示 y_1，y_2，…，y_n 的方差. 当 $y_1=x_1+b$，$y_2=x_2+b$，…，$y_n=x_n+b$ 时，验证 $s_y^2=s_x^2$.

习题 3

学而时习之

1. 下面的数据是 1900—1936 年奥林匹克男子跳高比赛金牌获得者的跳跃高度（单位：cm）. 计算均值、方差和标准差.

年　代	高　度	年　代	高　度
1900	190.0	1904	190.3
1908	190.5	1912	193.0
1920	193.5	1924	198.1
1928	194.1	1932	197.1
1936	202.9		

2. 某连锁超市销售部收到甲乙两厂家送来的质地相同的白糖各 10 包，测量后得到甲乙两厂家白糖的重量（单位：g）分别是：

甲厂	501	500	499	500	502	500	500	501	499	498
乙厂	497	501	500	502	499	501	503	500	500	497

销售部应当销售哪家的白糖？

3. 某公司希望能为飞机制造公司提供零部件，在向飞机制造公司推荐自己的生产能力时，应当重点明确以下哪些内容（　　）

(A) 所生产部件的平均规格符合标准

(B) 所生产部件的规格的方差不小于某个数

(C) 所生产部件的规格的方差不大于某个数

(D) 能够按时供货

4. 对一本书进行校稿前，抽查了其中的 21 页. 将排版时输入错误的情况总结如下：

输入错误数	1	7	4	0	11	6	2
出现总页数	3	6	3	2	2	1	4

(1) 计算每页的平均输入错误数；

(2) 计算样本方差；

(3) 计算样本标准差.

温故而知新

5. 当观测数据 x_1，x_2，…，x_n 的样本方差 $s^2=0$，证明所有的 x_i 相同.

6. 当数据 x_1，x_2，…，x_n 同时扩大 a 倍，证明方差扩大 a^2 倍.

7. 当数据 x_1，x_2，…，x_n 同时扩大 a 倍，证明标准差扩大 $|a|$ 倍.

12.2　抽样调查方法

在日常生活中人们总是自觉或不自觉地应用抽样方法，例如在市场上买花生或瓜子时总要先抓几颗看看是否饱满，干燥. 在厨房做饭

的过程中经常要取一点尝尝咸淡.

在考察锅里汤的味道时，没有必要把汤喝完，只要把汤搅拌均匀，从中品尝一勺就可以了．注意无论这锅汤有多少，只要一勺就够了．这就是窥一斑而知全豹.

记住上面的例子是大有好处的，因为它提供了抽样调查方法的最重要信息.

在刚加盐的地方舀出的汤作样本，你会作出汤太咸了的错误结论.

第一，把汤“搅拌均匀”是说明抽样的随机性，没有抽样的随机性，样本就不能很好地反映总体的情况.

第二，品尝一勺指出了选取的样本量不能太少，也不必太大．太少了不足以品出味道，品尝一大碗也没有必要.

抽样调查是相对于普查而言的．其含义是从总体中按一定的方式抽出样本进行考察，然后用样本的情况来推断总体的情况.

第三，“无论这锅汤有多少，只要一勺就够了”．这里体现出抽样调查的如下基本性质：总体个数增大时，样本量不必跟着增大.

抽样调查的必要性

在评价 1 000 台同型号的微波炉的平均工作寿命 μ 时，预备从中抽取 n 台进行工作寿命的测量试验，用这 n 台微波炉的平均工作寿命估计总体的平均工作寿命 μ.

这里，总体是 1 000 台微波炉的工作寿命，样本量是 n，被选中的微波炉的工作寿命构成样本．样本平均 $\bar{x}$ 是总体均值 μ 的估计.

根据喝汤的经验，没有必要调查很多微波炉的工作寿命.

在正确抽样的前提下，样本量越大，$\bar{x}$ 越接近总体均值 μ．但是，较大的样本量造成的损失很大．因为这 n 台微波炉做完寿命试验后就报废了．在本问题中要想得到真正的总体均值 μ 是不可能的，除非把这1 000台微波炉都拿来做工作寿命试验，报废掉这1 000台微波炉.

窥一斑而知全豹.

在很多实际问题中，采用抽样的方法来确定总体性质不仅是必要的，也是必须的.

12.2.1 随机抽样

在 12.1.2 的问题中，通过选取和测量 50 个同学的身高，得到了

总体平均身高 μ 的估计 $\bar{x}=161.86$（cm）．结果公布后，同学们普遍反映估计值偏低．其原因是什么呢？

造成估计值偏低的原因是抽样方案设计不合理．

原因在于女同学在选择调查对象时更倾向于，或更方便选择到女同学，所以 50 个同学的样本中女生身高占了绝大多数．这样就解释了估计值偏低的原因．

如何设计抽样方案才能得到满意的估计值呢？

在对总体的情况不清楚的时候，最好的抽样方案应当将总体中的个体一视同仁：每个个体被抽中的机会相同．

如果请 5 位男同学作抽样调查，就有可能得出估计值偏高的结果．

如果总体中的每个个体都有相同的机会被抽中，就称这样的抽样方法为**随机抽样**方法．

人们经常用“**任取**”，“**随机抽取**”或“**等可能抽取**”等来表示随机抽样．

例　口袋中有质地相同的小球 10 个，分 3 种颜色．从中无放回地随机抽取 1 个，共抽取 n（$\leqslant 10$）个，这种抽样的方法被称为无放回地随机抽样．从袋中每次随机抽取一球记录颜色后放回，共抽取 n 次，这样的抽样方法被称为有放回地随机抽样．这两种随机抽样有什么区别吗？

解　无放回随机抽样下，同一个小球不会被抽中两次．而有放回地随机抽样下，同一个小球可能被抽中多次．当样本量 $n=10$，采用无放回随机抽样就可以完全了解袋中小球的颜色分布情况，采用有放回地随机抽样还不能对袋中小球的颜色分布作出准确判断．

随机抽样又分为无放回地随机抽样和有放回地随机抽样．无放回地随机抽样指在总体中抽出一个个体后，下次在余下的个体中再进行随机抽样．有放回地随机抽样指抽出一个个体，记录下抽到的结果后放回，摇匀后再进行下一次随机抽样．

简单随机抽样指无放回地随机抽样．

简单随机样本指简单随机抽样得到的样本．

在没有特殊声明时，**所有的随机抽样都是指简单随机抽样**．

试验和理论都证明：在随机抽样下，样本均值 $\bar{x}$ 是总体均值 μ 很好的估计，样本标准差 s 是总体标准差 σ 很好的估计．在样本量不大

在相同的总体中和相同的样本量下，简单随机抽样得到的结果比有放回的随机抽样得到的结果要好．但是当总体的数量很大，样本量相对总体的数量又很小时，这两种抽样方法得到的结果是相近的．

时，增加样本量可以比较好地提高估计的精确度．

在 12.1.2 的问题中，实现简单随机抽样的方法是先将 500 个同学从 1 到 500 进行编号，然后将 500 张由 1 到 500 编号的小纸片放入一个大纸箱进行充分地摇匀．最后从纸箱中无放回地抽取 50 张纸片．纸片上的号码就是被选中的同学的号码．纸片上的这 50 个数被称为**随机数**（random number）．

摇匀后，一次取出 50 个和无放回地依次抽取 50 个的效果是相同的：都是没有重复的随机抽样．

随机数可以利用计算机产生．下面是用计算机在 1 至 500 中随机抽取的 50 个随机数．

50 个随机数

476	116	304	243	446	382	229	17	411	223
308	396	461	370	89	203	468	459	206	447
29	177	407	5	95	70	102	109	302	137
100	8	374	224	466	233	210	424	263	106
337	420	10	341	190	416	252	355	215	153

现在我们继续解决 12.1.2 中的问题．将 A 中学高一年级的 500 个同学从 1 到 500 进行编号，按照上述随机数表中的号码选取出 50 个同学，在 9 月的某一天测量他们的身高如下（单位：cm）：

165	165	169	162	165	163	165	177	162	156
169	159	166	161	170	170	164	151	167	177
173	155	165	161	159	157	168	163	166	163
163	168	171	181	156	159	173	163	156	168
177	171	180	169	179	163	158	164	174	158

用这 50 个观测数据计算出的样本均值是

$$\bar{x}=165.68\ (\text{cm}).$$

于是，高一年级同学平均身高的估计是 165.68 cm.

由于这次抽样是通过随机抽样完成的，因而避免了有偏的结果．这次抽样调查的结果得到了同学们的认可．

关于历史上不采取正确的抽样方案而导致调查结论严重失真的教训，可参阅本小节的阅读与思考．

练 习

1. 调查 1 000 只日光灯管的平均使用寿命时，随机抽样方法是有放回地随机抽样吗？
2. 调查长沙出租司机的月平均收入时，在街面上进行随机抽样调查，得到的样本是简单随机样本吗？

习题 4

1. 简述什么是简单随机抽样.
2. 简述什么是有放回地随机抽样.
3. 举例说明随机抽样方法中“随机”的必要性.
4. 在某一个海域对海豚的体重进行抽样调查时，应采用有放回地还是无放回地捕获抽样？
5. 在调查某个城市的家庭年平均收入时，能否只在该市的娱乐场所（如电影院、歌剧院、游乐场、健身馆等）进行随机抽样？原因是什么？能否只在该市的公共汽车站进行随机抽样？原因是什么？

阅读与思考

《文学摘要》的破产

1936 年是美国总统选举年．这年罗斯福（Roosevelt）任美国总统期满，参加第二届的连任竞选，对手是堪萨斯州州长兰登（Landon）．当时美国刚从经济大萧条中恢复过来，失业人数仍高达 900 多万，人们的经济收入下降了 1/3 后开始逐步回升．当时，观察家们普遍认为罗斯福会当选．而美国的《文学摘要》杂志的调查却预测兰登会以 57 % 对 43 % 的压倒优势获胜．《文学摘要》的预测是基于对 240 万选民的民意调查得出的．自 1916 年以来，在历届美国总统的选举中《文学摘要》都作了正确的预测．《文学摘要》的威信有力地支持着它的这次预测．

但是选举的结果是罗斯福以 62 % 对 38 % 的压倒优势获胜．此后不久《文学摘要》杂志就破产了．

要了解《文学摘要》预测失败的原因就必须检查他们的抽样调查方案．《文学摘要》是将问卷寄给了 1 000 万选民，这些选民的地址是在诸如电话簿、俱乐部会员名单等上查到的．

只收回了 240 万张问卷．

分析　1936 年只有大约 1/4 的家庭安装了电话．由于有钱人才更有可能安装家庭电话和参加俱乐部，所以《文学摘要》的调查方案漏掉了那些不属于俱乐部的穷人和没有安装电话的穷人，这就导致了调查结果有排除穷人的偏向．

在 1936 年，由于经济开始好转，穷人普遍有赞同罗斯福当选的倾向，富人有赞同兰登当选的倾向．《文学摘要》的调查结果更多地代表了富人的意愿，导致了预测的失败．

评论　抽样的方案应当公平地对待每一位选民和每一个群体，以便得到选民的真实情况．将哪一个群体排除在外的抽样方案都会

导致有偏的样本，从而导致错误的结论.

同一年，刚刚成立的盖洛普调查公司正确地预测了罗斯福获胜. 以后，盖洛普公司作过多次美国总统大选的民意调查. 由于采取了正确的抽样设计方案，在调查人数不是很多的情况下，预测的结果都是成功的.

12.2.2　调查问卷的设计

在随机抽样的前提下，具体抽样调查的实施方式也会影响调查的结果.

例　（敏感问题调查）某一个住宅区有2 060个家庭，调查人员已经拿到了所有住户的门牌号码. 在调查本住宅区存在家庭暴力的家庭比例时，因为经费的原因，只能调查 200 个家庭. 现在已经用随机抽样的方式抽出了 200 个门牌号码，问以下哪种调查方案较好？

（A）派调查员用询问记录的方式登门调查这 200 个家庭；

（B）登门调查这 200 个家庭前先准备 200 张问卷和若干支笔. 只需被调查者在下面的匿名问卷上打勾，然后请被调查者自己将问卷放入调查员的书包.

> 问卷：请选择［有家庭暴力］　［无家庭暴力］
>
> 我们承诺没有人知道你的回答是什么.

（C）采用和（B）相同的调查方案，但是告诉被调查者要调查 200 个家庭，而且让他们将答卷折叠后投入随身携带的封闭投票箱. 全部调查完毕后再开箱统计；

（D）将（B）中的 200 张问卷投入抽中的 200 个家庭的信箱，请他们在规定的时间内将答卷放入指定的投票箱.

解　由于家庭暴力是不光彩的家庭隐私，所以调查时应当让被调查者知道他的回答是得到严格保密的. 只有这样，调查才有可能获得事实真相. 按照这个原则，方案（C）是其中较好的方案.

方案（A）忽略了被调查者的感受，容易得到有偏的结果.

方案（B）记得带笔方便了调查，但是忽略了被调查者因不信任而不回答事实真相.

方案（D）的缺点是只能收回少量的问卷.

在抽样调查中，调查的方式方法也是非常重要的. 无论是当面调查还是问卷调查都应当做到以下两点：

（1）提问的内容要简单明确. 提问太长，会给回答带来困难和引

起被调查者的反感，不利于得到正确的回答.

（2）用词要确切，通俗易懂，有礼貌和不用引导词语.

例如：

“您用什么牌的牙膏?”时间范围不明确，应改为“您现在用什么牌的牙膏?”

“很多人都要买汽车，您呢?”带有引导性，应改为“您最近打算买汽车吗?”

练习

某校高中一年级有 6 个班，每班有 43 个同学，其中一个班是特长班. 该年级的全体同学已经将各自的学号写在规格相同的小纸条上，放入一个大纸箱中摇匀. 现在校长要通过抽样的方法调查该年级学生参加课外体育锻炼的情况，规定只对 36 个同学进行详细了解. 以下抽样方法正确的是（　　）

（A）从纸箱中无放回地随机取出 36 个学号，选取这 36 个学号的同学

（B）各班的班主任在自己的班上点出 6 个同学

（C）从纸箱中有放回地随机抽取 36 个学号，选取这 36 个学号的同学

（D）从特长班中用随机抽样的方法选取 36 个同学

习题 5

学而时习之

1. 用随机抽样的方法，在你的语文书中抽查 10 页. 回答以下问题.

（1）你的抽样是如何进行的，采用的是简单随机抽样还是有放回的随机抽样?

（2）这 10 页中，平均每页有多少个句号?

（3）你对语文书平均每页句号个数的估计是多少?

（4）如果另外的同学抽查了 20 页，你认为谁的估计更准确.

2.（数学实践）在抽样调查本校同学的手机个人拥有率、家庭汽车拥有率和每天完成作业所用的时间时，规定样本量为 $n=100$. 请同学设计一个合理的调查方案和一份调查问卷（参考 P. 76 例题中的问卷），并具体实施一次抽样调查工作.

12.2.3 分层抽样和系统抽样

分层抽样

要了解一盘菜炒得好吃不好吃，一般只要随机品尝几口就可以下结论了. 没有必要等到把菜吃完再作结论.

但是如果品尝的是西红柿炒鸡蛋，你进行随机抽样品尝就容易只品尝到西红柿或只品尝到鸡蛋，这对于你作出正确的判断是不利的. 你应当随机品尝一下西红柿，再随机品尝一下鸡蛋，然后进行综合评价. 这种品尝方法就是分层抽样方法.

例 1 某市进行家庭年收入调查时，分别对城镇家庭和农村家庭进行调查. 在全部城镇的85 679户中无放回随机抽取了 350 户，在全部农村的275 692户中无放回随机抽取了 360 户. 调查结果如下：

城镇家庭年平均收入是35 612元，

农村家庭年平均收入是5 623元.

试计算该市家庭年平均收入.

解 这里遇到了两个分总体 A_1 和 A_2，第一个分总体 A_1 是所有城镇家庭的年收入，第二个分总体 A_2 是所有农村家庭的年收入. 用 A 表示该市所有家庭的年收入时，总体 A 是两个分总体 A_1 和 A_2 的并.

用 $\overline{x}_1$ 表示来自总体 A_1 的样本均值，用 $\overline{x}_2$ 表示来自总体 A_2 的样本均值，则

$$\overline{x}_1=35\ 612, \qquad \overline{x}_2=5\ 623.$$

A_1 在 A 中所占的比例是

$$W_1=\frac{85\ 679}{85\ 679+275\ 692}=0.237\ 1.$$

A_2 在 A 中所占的比例是

$$W_2=\frac{275\ 692}{85\ 679+275\ 692}=0.762\ 9.$$

A 的总体均值 μ 的估计是

$$\begin{aligned}\overline{X}&=W_1\bar{x}_1+W_2\bar{x}_2\\&=0.237\ 1\times 35\ 612+0.762\ 9\times 5\ 623=12\ 733\ (\text{元}).\end{aligned}$$

于是该市平均年家庭收入的估计是 12 733 元.

分层抽样就是把总体 A 分成 L 个互不相交子总体：

$$A=A_1+A_2+\cdots+A_L.$$

称这些子总体为层，称 A_i 为第 i 层. 然后在每层中独立地进行简单随机抽样.

用 N 表示总体 A 的个体总数，用 N_i 表示第 i 层的个体总数时，有

$$N=N_1+N_2+\cdots+N_L.$$

我们称

$$W_i=\frac{N_i}{N},\ (i=1,\ 2,\ \cdots,\ L)$$

为第 i 层的**层权**（weight）.

用 μ 表示 A 的总体均值. 对 $i=1$，2，…，L，用 $\bar{x}_i$ 表示从第 i 层抽出样本的样本均值. 我们称

$$\overline{X}=W_1\bar{x}_1+W_2\bar{x}_2+\cdots+W_L\bar{x}_L$$

是总体均值 μ 的**简单估计**.

分层抽样是一种常用的抽样方法，有如下的特点：

（1）分层抽样在获得总体均值估计的同时，也得到各层的均值估计. 在例 1 中，不但得到了 A 的均值估计，还得到了 A_1 和 A_2 的均值估计.

（2）将差别不大的个体分在同一层，使得分层抽样得到的样本更具有代表性，从而提高估计的准确度.

（3）抽样调查的实施更加方便，调查数据的收集、处理也更加方便.

系统抽样方法

例 2 在调查某居民住宅区的 999 个住户对住宅区的环境满意程度时，是按照 1∶14 的比例进行抽样调查，试计算样本均值.

解 先将这 999 户按门牌号码的顺序依次编号，每个号对应一户的门牌号码.

1	2	3	4	5	6	7	…	13	14
15	16	17	18	19	20	21	…	27	28
29	30	31	32	33	34	35	…	41	42
…	…	…	…	…	…	…	…	…	…
981	982	983	984	985	986	987	…	993	994
995	996	997	998	999					

在 1～14 中随机抽取一个数字，如果抽到 7，就调查排在第 7 列的所有家庭，请这些家庭对小区环境的满意程度打分. 分数分为 1，2，3，4，5 级. 第 7 列有 71 户，所以样本量 $n=71$. 这 71 户的平均分是样本均值. 用样本均值作为全体住户对小区环境的平均分的估计.

用 x_i 表示这 71 户中第 i 户的打分，样本均值是

$$\overline{x}=\frac{x_1+x_2+\cdots+x_{71}}{71}.$$

我们称上面的抽样方法为系统抽样法.

如果总体中的个体按一定的方式排列，在规定的范围内随机抽取一个个体，然后按照制定好的规则确定其他个体的抽样方法称为**系统抽样方法**（systematic sampling method）.

最简单的系统抽样方法是取得一个个体后，按相同的间隔抽取其他个体.

系统抽样方法的主要优点是实施简单，只需要先随机抽取第一个个体，以后按规定抽取就可以了. 系统抽样法不像随机抽样方法，随机抽样方法每次都要随机抽取个体.

练习

A 中学高一年级的 500 名同学中有 218 名女生，在调查全年级同学的平均身高时，预备抽样调查 50 个同学. 请你做以下工作，并回答以下问题.

（1）设计一个合理的分层抽样方案.

（2）你的设计中，第一和第二层分别是什么？

（3）分层抽样是否在得到全年级同学平均身高的估计时，还分别得到了男生和女生的平均身高的估计.

习题 6

学而时习之

1. 调查你使用的语文书每页平均有多少个“。”，调查的比例是全书的 1/10 页.

（1）设计一个系统抽样方法；

（2）具体实施你的系统抽样方法，写出调查的样本，样本量；

（3）计算样本均值；

（4）你估计全书每页平均有多少个“。”；

（5）把你的结果和其他同学的结果进行比较，对比较的结果给出简单的分析.

2. 调查全班 49 个同学的平均身高时，决定采用系统抽样方法抽取 14 个样本作平均. 请 49 个同学按高矮顺序排列. 身高情况如下（单位：cm）：

150	153	155	155	155	156	156
157	157	158	158	158	159	159
160	160	160	160	160	161	161
161	162	162	162	162	163	163
163	164	164	164	164	165	165
165	165	165	165	166	166	166
166	167	167	168	168	170	172

应当采用抽样方法（　　）避免抽样偏差.

（A）前两行的样本平均

（B）后两行的样本平均

（C）前两列的样本平均

（D）两个对角线上数据的样本平均

3. 计算问题 2 中的总体平均和你所选用方法的样本平均.

12.3 用样本分布估计总体分布

大量的原始数据如果不经过有效的分析、整理，并通过适当的形式表示出来，就好比一堆没有经过冶炼的矿物，没有什么用途.

分析整理数据的方法之一是用图表把它们表达出来. 图表中包含的信息极多，因为数据中的大量信息都可以概括在图表内. 图表使人一目了然. 一幅图或一张表有时候往往胜过语言的表述.

无论是从抽样调查，还是从科学试验，工农业生产中得到的数据，在统计学中都被称为观测数据或样本. 观测数据也简称为**数据**，数据的个数被称为样本量.

在实际问题中，样本量往往是比较大的，这时数据中的主要信息隐藏在背后. 要从数据中得到这些信息，必须对观测数据进行整理. 下面是几种常用的数据整理方法.

12.3.1 频率分布表

当样本量是 n 的观测数据中有 n_i 个 y_i 时，我们称

$$f_i=\frac{n_i}{n}$$

是 y_i 出现的**频率** (frequency)，简称为 y_i 的频率. 例如数据

2，2，2，2，3，3，3，5，5，5

中，2 的频率是 4/10=0.4，3 的频率是3/10=0.3，5 的频率是 3/10 =0.3. 频率也可以用百分数表示. 在上面的例子中，2 的频率是 40 %，3 的频率是 30 %，5 的频率是 30 %.

案例 自 1500 年至 1931 年的 $n=432$ 年间，比较重要的战争(简称为战争)在全世界共发生了 299 次. 以每年为一个时间段的记录如下：

表 12.1

爆发的战争数 i	爆发 i 次战争的年数 n_i	频率 $f_i=n_i/n$
0	223	51.6 %
1	142	32.9 %
2	48	11.1 %
3	15	3.5 %
4+	4	0.9 %
总　计	432	100 %

其中第一行的 0，223，51.6 % 表示在 432 年中有 223 年发生战争的次数是 0，发生的频率是

$$f_1=\frac{223}{432}\approx 51.6\ \%;$$

第二行的 1，142，32.9 % 表示在 432 年中有 142 年发生战争的次数是 1，发生的频率是

$$f_2=\frac{142}{432}\approx 32.9\ \%;$$

……

第五行表示在 432 年中有 4 年发生战争的次数是大于等于 4 次的，发生的频率是

$$f_5=\frac{4}{432}\approx 0.9\ \%.$$

我们称表 12.1 是观测数据的**频率分布表**（frequency distribution table）. 它简化了 432 年中有关战争爆发的 432 个观测数据，帮助我们更清楚地看到战争爆发的特征和规律.

制作频率分布表时，先将数据从小到大排列，然后将排列后的数据进行分段，相等的数据分在同一段内. 每段中的数据被称为一组数据，所以我们又把分段称为**分组**. 一般来讲，当样本量是 n，可以参照下面的经验公式将数据分成大约

$$K=1+4\lg n$$

段. 但是这里的经验公式只对分段起参考作用. 实际应用时，应当根据样本量的大小和数据的特点以及分析的要求灵活确定.

让我们通过例子学习频率表的制作方法.

例 下面是某城市公共图书馆在一年中通过随机抽样调查得到的 60 天的读者借书数，数据已经从小到大排列，制作频率分布表.

213	230	239	289	291	301	308	310	311	312
318	318	337	343	344	348	349	351	360	362
368	372	374	379	383	385	390	393	396	399
400	404	406	425	429	430	436	438	440	441
444	446	450	453	456	458	471	473	475	483

484　495　498　498　521　524　549　556　568　584

解　数据中的最小值是 213，最大值是 584．这 60 个数据就散布在闭区间 [213，584] 中．取一个略大的区间 (200，600]，它的端点都是整数．用经验公式计算出

$$K=1+4\lg n=1+4\lg 60\approx 8.$$

将 (200，600] 八等分，排在下表的第一列．计算出数据落入各段的个数 n_i，填入第二列．计算出数据落入各段的频率

$$f_1=\frac{3}{60}=5\ \%,\quad f_2=\frac{2}{60}\approx 3.3\ \%,\quad \cdots,\quad f_8=\frac{3}{60}=5\ \%,$$

依次填入第三列．最后将各列之和填入最后一行，得到频率分布表 12.2.

表 12.2

借出书数 i	发生次数 n_i	f_i＝发生频率
(200，250]	3	5 %
(250，300]	2	3.3 %
(300，350]	12	20 %
(350，400]	14	23.3 %
(400，450]	12	20 %
(450，500]	11	18.3 %
(500，550]	3	5 %
(550，600]	3	5 %
总　　计	60	99.9 %

由于计算频率时四舍五入引起计算误差，频率之和可能是 1 的近似.

从上述频率分布表可以方便地分析出以下结果：

有 8.3 % 的工作日借出的图书少于等于 300 册；

有 63.3 % 的工作日借出图书的数量在 301 至 450 册之间；

有 48.3 % 的工作日借出的图书在 400 册以上；

只有 10 % 的工作日借出的图书多于 500 册.

当总体是全年每个工作日的借书数量时，上述结果可以作为对总体的推测.

从上例可以总结出制作频率分布表的一般步骤如下：

第一步：将数据从小到大排列，将排列后的数据进行分段，相等的数据必须分在同一段内．每段中的数据被称为一组数据，所以我们

又把分段称为**分组**.

分段的多少应当适中．分段过多，数据过于分散，不利于看出数据的特征和规律；分段过少也不利于看到数据的特征和规律．当样本量是 n，可以参照经验公式将数据分成大约 $K=1+4\lg n$ 段.

也可以用 $K=\sqrt{n}$ 作经验公式.

第二步：决定各段的长短．在许多情况下，为了方便，除去第一和最后的两段，可以把其他各段的长度取作相同．还应当把各段的端点确定在便于记忆的数值上．为了达到以上目的，第一段的左端点可以比数据的最小值小一些，最后一段的右端点可以比数据的最大值大一些.

第三步：绘制频率分布表的第一列（参考上例）.

第四部：计算每段内数据的个数 n_i，填入表格的第二列.

第五步：计算数据落在第一段内的频率 f_i，填入表格的第三列.

第六步：将第 2，第 3 列之和填入最后一行.

说明：由于频率分布表的制作没有统一的数据分段方法，所以对相同的数据，同学们可以作出不同的频率分布表．但是好的频率分布表应当是简单明了的.

练　习

制作习题 6 第 2 题中 49 个同学身高的频率分布表.

习题 7

学而时习之

用随机抽样方法调查了某城市 50 辆公交车的营业额（单位：元），数据已从小到大排列.

259	294	295	297	300	300	300	301	301	302
303	306	308	309	311	314	315	315	321	323
327	328	331	334	336	339	339	339	347	348
350	350	352	355	359	359	361	363	370	376
377	383	388	389	390	396	404	410	410	411

（1）制作频率分布表；

（2）对频率分布表进行简单的分析（参考 P. 83 的例题分析）.

12.3.2 频率分布直方图

从文献记载上看，直方图在 1895 年由著名的英国统计学家**皮尔逊**（Pearson）作了描述，这可能是直方图的第一次使用. 他在为伦敦皇家协会发表的讲话中，当谈及 1885 年至 1886 年英格兰房地产估价的时候使用了直方图.

数据的频率分布图初步展示了数据分布的一些规律. 如果用图形来表示频率分布就会更加形象和直观. 显示数据频率分布的图形有频率分布直方图和茎叶图.

有了数据的频率分布图，很容易作出频率分布的直方图.

将观测数据按照制作频率分布表的方法进行分段，计算出数据落入各段的频率 f_i，将各段的端点画在直角坐标系中的横坐标上. 用 f_i 作为纵坐标的高，就得到了由相连接长方形构成的图形. 我们把所得到的图形称为数据的频率分布直方图，简称为**直方图**(histogram).

例 绘制 P. 83 例题中图书馆借出图书数据的频率分布直方图.

解 在横坐标上标出所有的数据分段的端点，

$$200,\ 250,\ \cdots,\ 550,\ 600.$$

在区间 [200，250] 上绘制以频率 0.05 为高的矩形；

在区间 [250，300] 上绘制以频率 0.033 为高的矩形；

……

在区间 [550，600] 上绘制以频率 0.05 为高的矩形.

就得到了需要的频率分布直方图. 如图 12 - 1.

从频率分布直方图可以更直观地看到图书馆每日借出图书册数的分布情况.

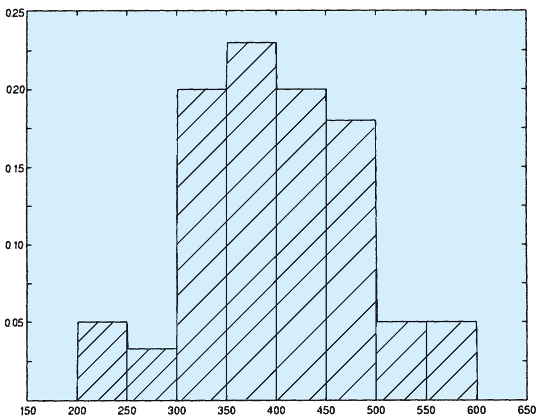


图 12－1　频率表 12.2 的直方图

练　习

制作习题 6 第 2 题中 49 个同学身高的频率分布直方图.

习题 8

学而时习之

下面是 1997 年至 2000 年广州的月降水量（单位：mm，数据摘自中国气象年鉴），绘制频率分布直方图.

1997 年	66	106	60	197	166	469	248	282	204	143	10	49
1998 年	73	112	41	245	306	370	223	121	165	48	22	10
1999 年	34	0	62	117	152	152	176	496	273	40	26	54
2000 年	11	30	28	419	203	197	288.9	185	57	304	34	43

12.3.3 频率折线图

用 d_1，d_2，…，d_k，分别表示频率分布直方图中各矩形上边的中点，在直方图的左边延长出一个分段，分段的中点用 d_0 表示. 在直方图的右边也延长出一个分段，分段的中点用 d_{k+1} 表示.

用直线连接 d_0，d_1，…，d_{k+1} 就得到了一条折线，这条折线叫作频率折线图. 频率折线图也反映出数据频率分布的规律.

图 12－2 是 P. 83 的例题中图书馆借出图书数目的频率折线图.

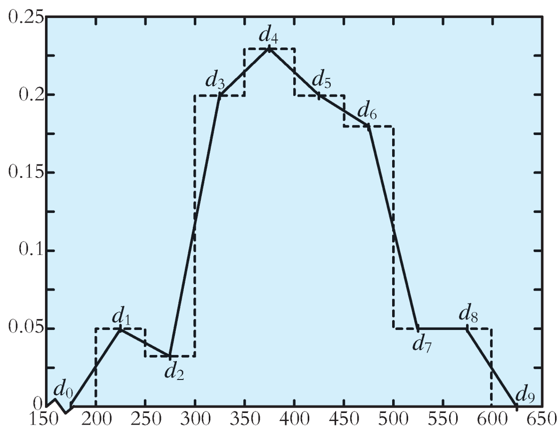


图 12－2　频率表 12.2 的频率折线图

案例　（选择性繁殖问题）为了研究老鼠的智力能否遗传，柏克莱（Berkeley）大学教授做了以下的试验. 分别让 142 只老鼠走相同的迷宫，每只老鼠走 19 次. 老鼠犯错误的次数就是走不出迷宫的次数. 我们把犯错误少的老鼠称为伶俐老鼠. 记录每只老鼠犯错误的次数，得到 142 个数据. 这 142 个数据的频率折线图如图 12－3.

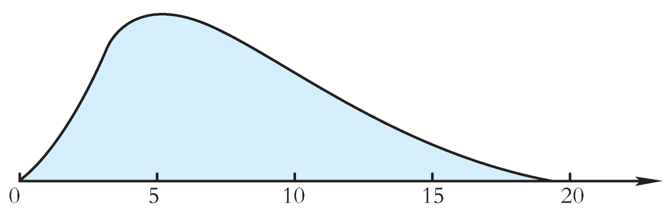


图 12－3

试验后把伶俐的老鼠放在一起，让它们进行繁殖，把不伶俐的老鼠放在一起进行繁殖. 繁殖 7 代之后，得到伶俐组的后代 85 只，非

伶俐组的后代 68 只．让这两组老鼠再走相同的迷宫，每只走 19 次．得到各组老鼠犯错误的次数后，为两组数据作出的频率折线图如图 12－4．

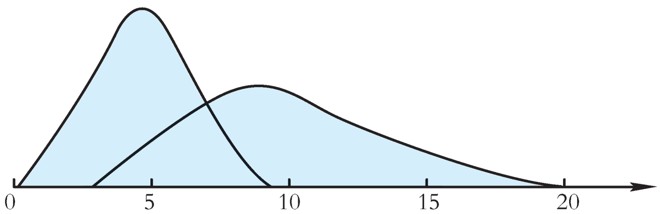


图 12－4

左面的折线图是伶俐组后代的频率折线图，右面的折线图是非伶俐组后代的频率折线图．这两条折线图有明显的差异．伶俐组的后代犯错误的次数明显地少，说明老鼠走迷宫的能力是具有遗传性的．

练　习

制作习题 6 第 2 题中 49 个同学身高的频率折线图．

习题 9

学而时习之

对于习题 8 中 1997 年至 2000 年的广州月降水量数据，绘制频率折线图．

12.3.4　数据茎叶图

直方图主要用于展示分段数据的频率分布，对于没有分段的观测

数据还可以用数据的茎叶图展示它的特性.

数据的**茎叶图**（stemplot）由“茎”和“叶”两部分组成，在制作茎叶图的时候要先确定数据的“茎”和“叶”. 从数据的茎叶图可以看出数据的分布形状及数据是否对称，是否集中等分布特性.

我们通过举例说明茎叶图的制作方法.

例 1 下面是上海市 2004 年 7 月 11 日至 2004 年 8 月 1 日空气中可吸入颗粒物的监测数据. 为这批数据制作茎叶图.

85 85 66 71 62 52 55 59 52 62 59

70 80 96 97 94 62 51 57 67 96 93

解 先将数据从小到大排列得到：

51 52 52 55 57 59 59 62 62 62 66

67 70 71 80 85 85 93 94 96 96 97

数据的十位上的数是 5，6，7，8，9，把它们叫作“茎”，排列下面茎叶图的第一列；

茎 5 后面的个位数分别是 1，2，2，5，7，9，9，把它们叫作茎 5 的“叶”，排在茎 5 的后面；

按相同的方法把茎 6 的叶 2，2，2，6，7 排在茎 6 的右边；

……

把茎 9 的叶 3，4，6，6，7 排在茎 9 的右边.

就得到了如下的茎叶图.

树茎	树叶
5	1225799
6	22267
7	01
8	055
9	34667

优：可吸入颗粒物在 0～50 之间.

良：可吸入颗粒物在 51～100 之间.

轻度污染：可吸入颗粒物在 101 ～ 150 之间.

从茎叶图中看出，尽管这 22 天中可吸入颗粒物都是处于良的水平，但是有较多的时间接近于优，也有较多的时间接近于轻微污染.

在同一个茎叶图中还可以表现两组数据的分布情况，这样做有利于对这两组数据进行比较. 我们称表示两组数据的茎叶图为**双茎叶图**.

例 2 下面是上海市 2004 年 7 月 11 日至 2004 年 8 月 1 日空气中二氧化硫和二氧化氮的监测数据. 为这两组数据制作一个双茎叶图，并进行比较.

二氧化硫数据：55，62，54，71，60，51，55，56，51，58，61，62，69，73，72，69，58，42，42，65，77，73.

二氧化氮数据：38，37，30，39，31，19，22，22，18，26，25，31，38，44，42，35，22，19，22，37，50，38

解 先将两组数据分别从小到大排列，得到：

二氧化硫数据：42，42，51，51，54，55，55，56，58，58，60，61，62，62，65，69，69，71，72，73，73，77.

二氧化氮数据：18，19，19，22，22，22，22，25，26，30，31，31，35，37，37，38，38，38，39，42，44，50.

这两组数据都是十位数，选用十位上的数作“茎”，排在双茎叶图的中间一列，它们是 1，2，…，7.

然后将二氧化硫的各位数作为“叶”，依次排放在相应的茎的左边. 例如，从数据 42，42 得到茎 4 的叶 22，排在 4 的左边；

从数据 51，51，54，55，55，56，58，58 得到茎 5 的叶 11455688，将它们从大到小排在 5 的左边得到 88655411；

……

把茎 7 的叶 12337 从大到小排在 7 的左边；

再按照例 1 的方法把二氧化氮数据排在“茎”的右边.

就得到了两组数据的双茎叶图：

二氧化硫 树叶	树茎	二氧化氮 树叶
	1	899
	2	222256
	3	0115778889
22	4	24
88655411	5	0
9952210	6	
73321	7	

从上述茎叶图可以看出，二氧化硫的空气质量指标比二氧化氮的空气质量指标要差很多. 二氧化硫的空气质量指标基本都处在良好的水平，而二氧化氮的空气质量指标都处在优的水平.

数据茎叶图的优点是显示了数据的每个信息，从茎叶图中可以直观地看到数据的分布情况. 但是数据量很大时，茎叶图的效果就不好了，因为这时的茎叶图会很长或很宽.

多知道一点

数据的茎叶图

茎叶图的茎也可以是两位或三位数.

例 制作以下两组数据的双茎叶图.

数据 1：112，113，115，123，126，127，132，137，138，139，142，143，156，158，159，161，164，165，165.

数据 2：123，124，126，134，135，135，137，142，143，146，147，149，152，158，159，162，163，164.

解 数据都是三位数，我们以前两位数作茎，它们分别是 11，12，13，14，15，16. 按照例 2 的方法可以作出双茎叶图如下：

数据 1 树叶	树茎	数据 2 树叶
532	11	
763	12	346
9872	13	4557
32	14	23679
986	15	289
5551	16	234

练习

分别制作下面的习题中数学、物理、语文考试成绩的茎叶图.

习题 10

学而时习之

甲班的期中考试成绩排列如下：

数学：65，65，67，68，70，71，75，76，77，79，80，82，83，83，83，84，84，85，86，86，87，88，88，88，91，93，93，93，93，93，94，95，97，99，99，99，100，100，100，100.

物理：57，60，62，64，67，67，70，72，73，74，74，74，77，77，78，78，79，80，80，81，81，81，82，82，84，84，84，84，86，87，88，93，95，96，96，98，99，99，100，100.

语文：62，67，67，70，70，70，71，72，72，73，74，75，75，76，76，77，78，78，78，78，79，79，80，80，80，80，81，82，82，82，83，83，84，84，86，88，89，90，93，95.

分别对数学和物理、物理和语文、数学和语文制作双茎叶图. 通过对茎叶图的观察回答以下问题.

（1）哪科平均成绩最好？

（2）哪科成绩分布最集中？

12.4 数据的相关性

在实际问题中，我们经常遇到有相关关系的变量．比如讲身高与体重的关系时，虽然身高不能确定体重，但总的来讲，身高者，体也重．

在考虑某一个特定地区居民的身高和体重的关系时，用 x 表示人的身高，用 y 表示体重，总体来讲，y 随着 x 增大一般也会增大．这时我们称 x 和 y 有**相关关系**．

在某地区的 12～30 岁居民中随机抽取了 10 个样本，用 x_i 和 y_i 分别表示第 i 个人的身高和体重，得到的数据如下：

身高（cm）	143	156	159	172	165	171	177	161	164	160
体重（kg）	41	49	61	79	68	69	74	69	68	54

数据 x_i 和 y_i 是成对出现的，所以用（x_i，y_i）表示第 i 个人的身高和体重．这时称数据对

$$(x_i, y_i), i=1,2,\cdots,10$$

为样本或观测数据．样本是直角坐标系中的 10 个点，将这 10 个点画在坐标系上得到的图称为观测数据的**散点图**（scatter diagram）．见图 12－5．

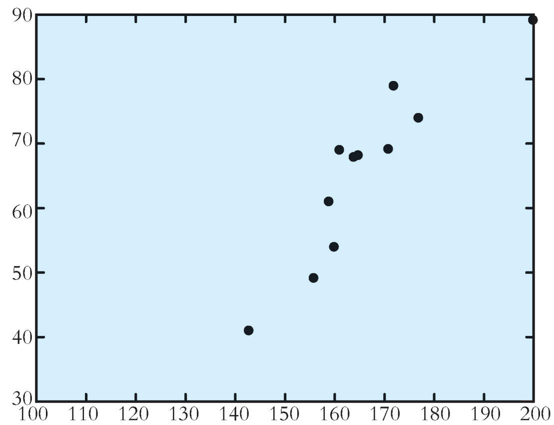


图 12－5　例中数据的散点图

上面，用（x，y）泛指总体中某个体的身高和体重时，我们把身高和体重的关系说成是 x 和 y 的关系．

12.4.1 相关性

无论是从抽样调查中得到的成对数据，还是从科学试验、工农业生产中得到的成对数据，在统计学中也都称为观测数据或样本，称数据对的个数为样本量.

样本量是 n 的成对观测数据是用

$$(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$$

表示的. 这里，对固定的 i，x_i 和 y_i 或是来自相同的个体，或是同一次试验的观测数据. 对 $i \neq j$，(x_i, y_i) 和 (x_j, y_j) 或是来自不同的个体，或是不同试验的观测数据.

在图 12－5 中，随着身高 x 的增加，体重 y 有明显的增加趋势. 这时称 x 和 y 是正相关的. 当数据

$$(x_1,\ y_1),\ (x_2,\ y_2),\ \cdots,\ (x_n,\ y_n)$$

十分明显地集中在一条上升的直线附近时，我们称 x 和 y 是高度正相关的.

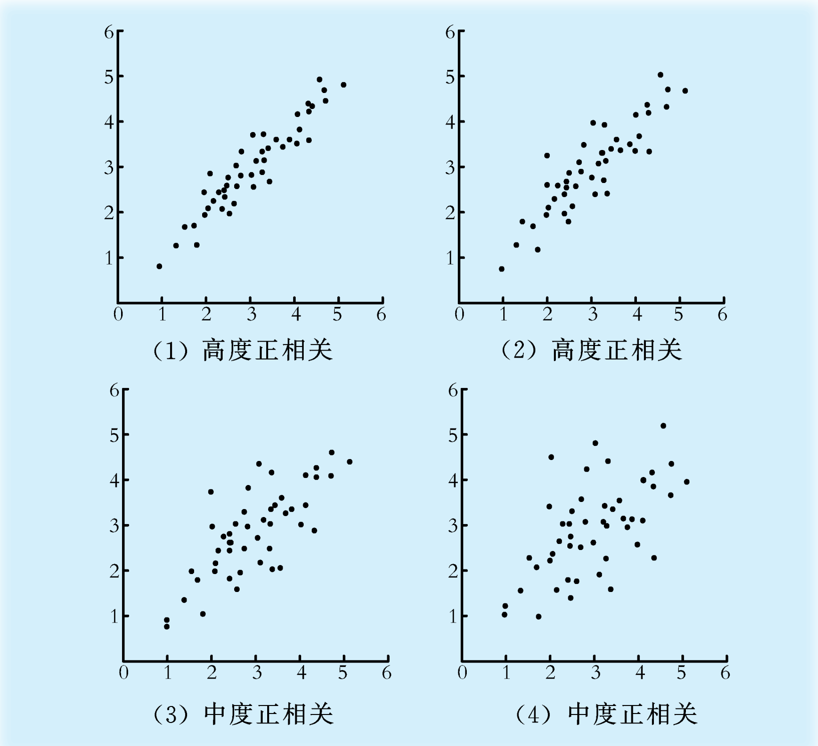


图 12－6

图 12－6 的(1)和(2)展示的数据是高度正相关的．当上述数据也分布在一条上升的直线附近，但集中的程度不十分明显时，我们称 x 和 y 是中度正相关的．图 12－6 中(3)和(4)展示的数据是中度正相关的．

当数据 (x_1, y_1)，(x_2, y_2)，…，(x_n, y_n) 十分明显地集中在一条下降的直线附近时，我们称 x 和 y 是高度负相关的．图 12－7 中(1)和(2)展示的数据是高度负相关的．图(3)和(4)展示的数据是中度负相关的．

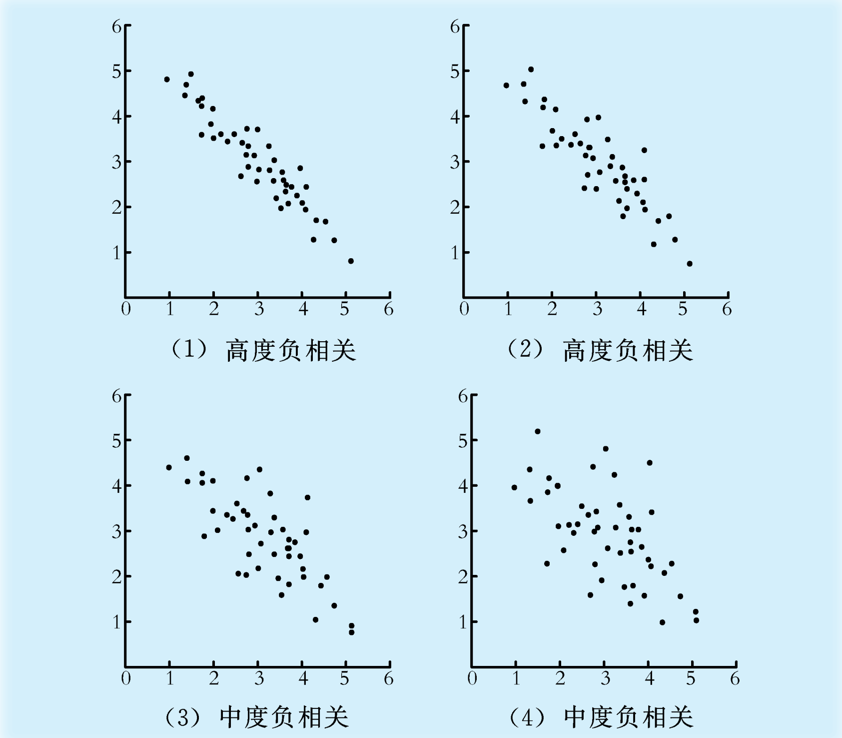


图 12－7

我们把有高度相关性或中度相关性的数据统一称为有相关性的数据．图 12－6 和图 12－7 中展示的数据都是具有相关性的数据，这时也称 x 和 y 是相关的．

练习

绘制以下数据的散点图．

年代 x	1975	1976	1977	1978	1979
比萨斜塔倾斜量 y	642	644	656	667	673

习题 11

2000 年的 3 月至 10 月北京和广州的月平均气温（单位:℃）记录如下.

月份	3	4	5	6	7	8	9	10
北京	8	15	20	27	30	26	22	13
广州	19	23	26	28	29	28	27	25

分别绘制北京和广州月平均气温的散点图.

12.4.2 回归直线

当$\{x_i\}$和$\{y_i\}$相关时，我们已经知道数据

$$(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$$

会分散在一条直线的附近. 我们将这条直线叫作回归直线，下面就寻找这条直线.

在直角坐标系中，两个点(x_1, y_1)，(x_2, y_2)可以决定一条直线.

$$l: y=\frac{y_2-y_1}{x_2-x_1}(x-x_1)+y_1, \text{当 } x_2 \neq x_1,$$

这时，两个点都在直线上，所以这两个点平均距离直线 l 最近.

给定三对观测数据，

$$(x_1, y_1), (x_2, y_2), (x_3, y_3)$$

当 x_1，x_2，x_3 不全相同，我们也求一条直线 l，使得以上三个点距离直线 l 平均最近.

用 $$l: y=bx+a$$

表示要求的直线，在平行于 y 轴的方向，作以上三点到直线 l 的连线，交点 A，B，C 的坐标见图 12－8.

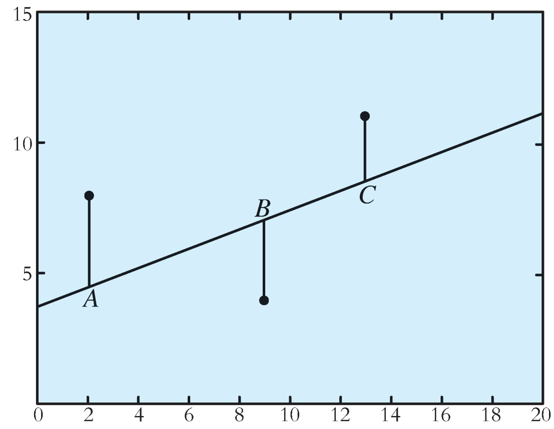


图 12－8

$$A:(x_1,bx_1+a),B:(x_2,bx_2+a),C:(x_3,bx_3+a).$$

三对观测数据和它们与直线 l 交点的距离分别是

$$|y_1-(bx_1+a)|,\ |y_2-(bx_2+a)|,\ |y_3-(bx_3+a)|.$$

我们用这三个距离的平方和

$$(y_1-bx_1-a)^2+(y_2-bx_2-a)^2+(y_3-bx_3-a)^2$$

衡量这三个观测数据远离直线 l 的程度. 如果 a，b 使得

$$Q(a,b)=(y_1-bx_1-a)^2+(y_2-bx_2-a)^2+(y_3-bx_3-a)^2$$

达到最小，就称直线 l：$y=bx+a$ 是回归直线.

不用 $|y_1-(bx_1+a)|+|y_2-(bx_2+a)|+\cdots+|y_3-(bx_3+a)|$ 的原因是数学上不好处理.

一般地，要为样本量是 n 的观测数据

$$(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)\quad（其中的 x_i 不全相同）$$

建立一条直线 l：$y=bx+a$，使之与观测数据平均最近时，也采用相同的方法. 沿平行于 y 轴的方向，点 $(x_i,\ y_i)$ 到它与 l 的交点的距离是

$$|y_i-(bx_i+a)|,\ i=1,2,\cdots,n.$$

我们用这些距离的平方和

$$Q(a,b)=(y_1-bx_1-a)^2+(y_2-bx_2-a)^2+\cdots+(y_n-bx_n-a)^2$$

衡量观测数据远离直线 l 的程度. 如果常数 a，b 使得 $Q(a,b)$ 达到最小，就称直线

$$l:\ y=bx+a$$

是$\{x_i\}$与$\{y_i\}$的**回归直线**.

得到了回归直线后，只要$\{x_i\}$与$\{y_i\}$高度正相关或高度负相关，对于新的 x，就可以用回归直线上的点 $y=bx+a$ 作为 y 的预测值.

用 $\bar{x}$ 和 $\bar{y}$ 分别表示$\{x_i\}$和$\{y_i\}$的样本均值，用 s_x^2 表示$\{x_i\}$的方差. 引入符号

$$s_{xy}=\frac{x_1y_1+x_2y_2+\cdots+x_ny_n}{n}-\bar{x}\bar{y}.$$

可以证明，只要 x_1，x_2，…，x_n 不全相同，回归直线中的

$$b=\frac{s_{xy}}{s_x^2}，\ a=\bar{y}-b\bar{x}.$$

例 1　下面是某冷饮部 8 天中出售的冷饮杯数和当天最高气温的记录数据.

为数据建立回归直线. 在同一个坐标系中画出数据的散点图和回归直线. 对最高气温是 30 ℃的一天，预测大约能出售多少杯冷饮.

气温（℃）	26	29	17	23	36	34	5	32
杯数	36	37	29	37	52	49	19	47

解　用 x 表示气温，用 y 表示杯数.

先计算出 $\bar{x}=25.25$，$\bar{y}=38.25$，再计算出

$$s_{xy}\approx95.44，\ s_x\approx9.59，\ s_y\approx10.28.$$

最后得到

$$b=\frac{s_{xy}}{s_x^2}\approx\frac{95.44}{9.59^2}\approx1.04;$$

$$a=\bar{y}-b\bar{x}\approx38.25-1.04\times25.25\approx12.$$

回归直线是 $y=1.04x+12$. 从数据的散点图看出卖出的冷饮杯数$\{x_i\}$和最高气温$\{y_i\}$是高度正相关的.

当 $x=30$ ℃，对 y 的预测值是

$$y=1.04\times30+12=43.2.$$

最高气温是 30 ℃时，大约可以出售 43 杯冷饮.

$\{x_i\}$和$\{y_i\}$的散点图和回归直线见图 12－9.

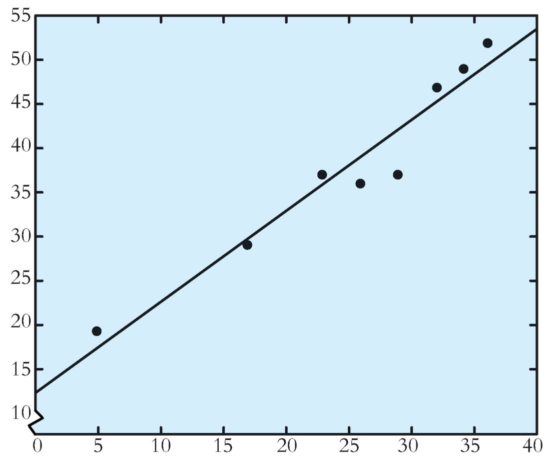


图 12－9

例 2 在测量一根新弹簧的劲度系数时，测得了如下的结果.

所挂重量（N）	1.00	2.00	3.00	5.00	7.00	9.00
弹簧长度（cm）	10.18	11.21	11.85	13.01	14.29	15.03

（1）建立回归直线；

（2）在同一个坐标系中绘制数据的散点图和回归直线.

解 用 x 表示所挂重量，用 y 表示弹簧长度. 按例 1 中的方法可以计算出

$$\bar{x}=4.50,\ \bar{y}=12.595,\ s_{xy}\approx4.74,\ s_x\approx2.81.$$

从数据的散点图 12－9 看出 $\{x_i\}$ 和 $\{y_i\}$ 是高度正相关的. 再计算

$$b=\frac{s_{xy}}{s_x^2}=\frac{4.74}{2.81^2}\approx0.6,$$

$$a=\bar{y}-b\bar{x}\approx12.595-0.6\times4.5\approx9.9.$$

于是，回归直线是

$$l:\ y=0.6x+9.9.$$

回归直线和数据的散点图见图 12－10.

我们知道弹簧的拉伸长度和所挂重量之间的关系服从胡克(Hooke)定律：

$$y=b_0x+a_0.$$

其中 x 是弹簧所挂重物的重量（N），y 是弹簧的全长. 这里的 a_0 和

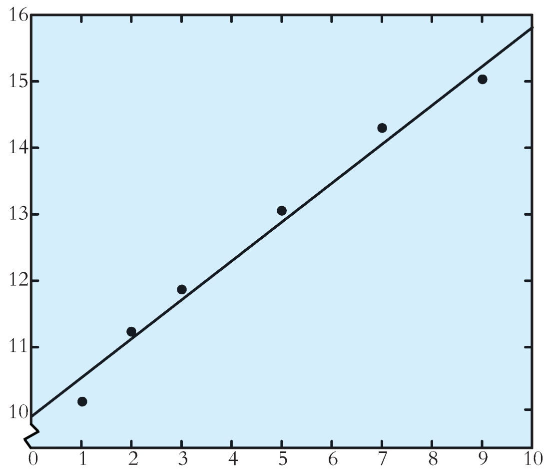


图 12－10

b_0 是未知的．所以 $b=0.6$ 和 $a=9.9$ 分别是 b_0 和 a_0 的估计．

由于 a_0，b_0 是弹簧本身固有的量，把它们称为参数(parameter)．由于 a，b 是通过求平方和 $Q(a,b)$ 的最小值得到的，所以我们称回归直线中的 a，b 分别是参数 a_0，b_0 的**最小二乘估计**．这里最小是求最小值的意思，二乘是平方和的意思．

练　习

建立 $\{x_i\}$ 与 $\{y_i\}$ 的回归直线：

年代 x	1984	1985	1986	1987
比萨斜塔倾斜量 y	717	725	742	757

习题 12

1．当 x_1，x_2，…，x_n 不全相同，证明 $(\bar{x}, \bar{y})$ 总在回归直线上．

2．以下数据来自对 5 个职工的随机抽样调查．x 表示月平均收入，y 表示用于购书和买报纸的月平均支出．(单位：元)

职工	1	2	3	4	5
x	1 200	1 400	1 700	1 800	2 100
y	12	11	14	18	21

(1) 对数据建立回归直线；

(2) 对月收入为 1 850 元和 1 500 元的职工，预测每月平均用于购书和买报的支出；

(3) 在同一个坐标系中绘制散点图和回归直线.

3. 以下是 4 个同学每天平均完成作业的时间和每天平均看电视时间的调查结果，调查在同一个班内用简单随机抽样方法完成.（单位：小时）

同学	1	2	3	4
做作业的时间 x	2	2.5	3.2	3
看电视的时间 y	1.5	0.6	0.1	0.3

(1) 为数据建立回归直线；

(2) 对于平均每天用 1.9，2.8 和 3.2 小时完成作业的同学，预测平均每天看电视的时间；

(3) 在同一个坐标系中绘制散点图和回归直线.

多知道一点

使用计算机或计算器作统计计算

MatLab 是一个应用广泛的计算机软件，功能强大，使用方便，易学易懂. MatLab 的各种版本大同小异，低版本的命令可以在高版本上应用. 本书的统计部分使用 MatLab 进行计算.

通过下面的例子可以初步学会使用 MatLab.

1. 计算 156，165，…，161 均值时，用下面的两个语句. 每个语句后面敲换行（Enter）. 括弧中的内容是语句的解释，不输入.

x=[156;165;…;161]；

M=mean(x) （计算 x 的均值）.

2. 将数据 156，165，…，161 从小到大排序时，用下面的语句：

x=[156;165;…161]

sort(x)

3. 计算 $n=50$ 个数据 60，65，65，…，98 的方差时，用下面的语句：

$n=50$

y=[60；65；65；…；98]

s2=Std(y)^2*(n−1)/n （计算 x 的方差，n 是样本量）.

4. 计算 $n=50$ 个数据 60，65，65，…，98 的标准差时，用下面的语句：

$n=50$

y=[60；65；65；…；98]

s1=Std(y)*sqrt((n−1)/n) （计算 y 的标准差）.

5. 计算回归直线中的最小二乘估计 a，b 时，用下面的语句. 每个语句后面敲换行.

x=[1；2；3；…；7；9]

y=[10.18；11.21；…；15.03]

ba=polyfit(x，y，1) （计算 b 和 a，输出的第一个数是 b，第二个数是 a）

b a

0.598 6 9.901 2 （得到的结果）.

通过下面例子学会使用计算器作统计计算.

1. 计算 3，5，8，12，19，21.5，32.3 的样本均值、样本标准差和样本方差.

MODE 2 （进入计算模式）

SHIFT SCL = （清空统计存储器）

3 DT 5 DT 8 DT 12 DT 19 DT 21.5 DT 32.3 DT

SHIFT $\bar{x}$ = 14.4 （计算的样本均值）

SHIFT $x\sigma_n$ = 9.688 137 076 （计算的样本标准差）

x^2 = 93.86 （计算的样本方差）

2. 计算回归直线中的参数 a，b，对 $x=19$ 预测 y.

x	10	15	20	25	30	35
y	103	105	110	115	120	124

MODE 3 1 （进入回归模式）

SHIFT SCL = （清空统计存储器）

10，103 DT 15，105 DT 20，110 DT 25，115 DT 30，120 DT 35，124 DT

SHIFT A = 92.904 761 9 （计算的 a）

SHIFT B = 0.885 714 285 （计算的 b）

19 SHIFT $\hat{y}$ 109.733 333 3 （对 $x=19$ 预测的 y）

数学实验

用计算机画回归直线和作统计计算

打开用“Z＋Z 超级画板”制作的课件“回归直线.zjz”（可以从本书配套光盘中复制或从网站 www.zplusz.org 下载），屏幕画面如图 12－11：

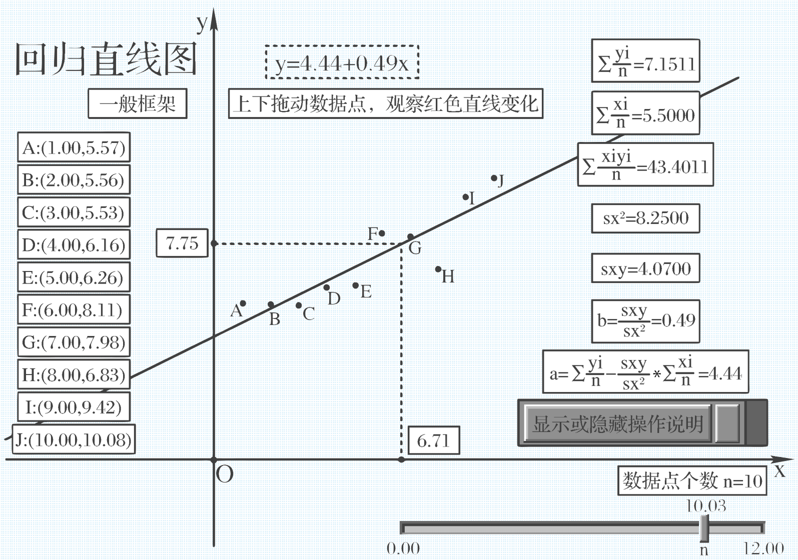


图 12－11

图上显示出有 10 个数据点的散点图和对应的回归直线．这些数据点可以上下拖动，回归直线随之变化．左边显示出每个点的坐标，即数据对；右边显示出为画出回归直线而必须作的统计计算结果．

鼠标单击右下方“显示或隐藏操作说明”按钮，仔细阅读出现的文本内容；根据操作说明，输入本章 12.4.2 节例 2 中的数据，并把点的个数调整为 6，就得到书中图 12－10 所示的回归直线．

单击上方的“下一页”图标，屏幕上出现课件的第二页，

如图 12－12，即例 2 的回归直线图．对照一下，和你作的是否相同？如果不同，检查自己的操作是否有误．

拖动横坐标轴上的红点，注意淡蓝色小框中数字的变化，想一想，这些数字的意义是什么？能利用这些数字预测弹簧所挂重量为 4.5 kg 时的长度吗？

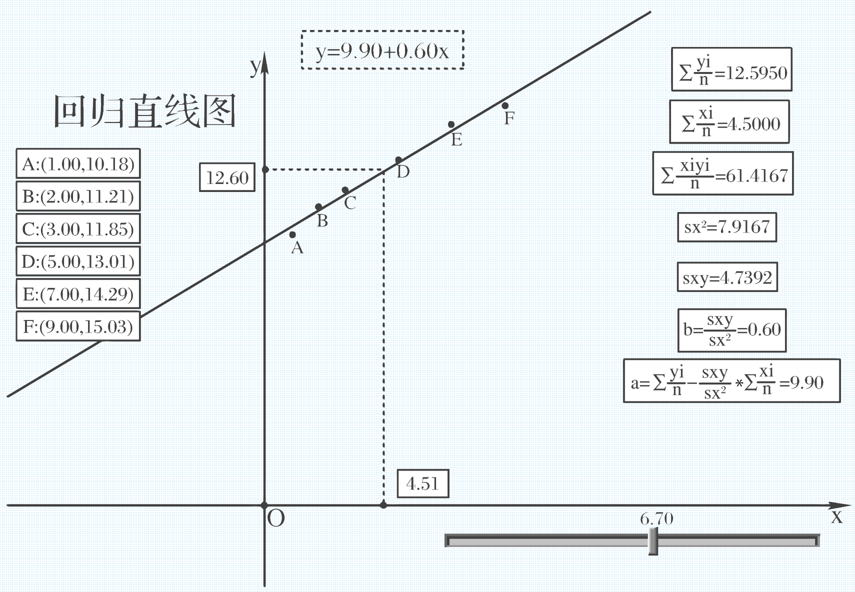


图 12－12

类似地，可作出本章 12.4.2 的例 1 的图，在图上预测最高气温为 30 ℃时出售冷饮的数量．

课件中设置了自动计算统计量的功能．为了熟悉统计量的计算，也可以应用前一章学到的算法知识，自己写程序计算统计量．

下面举例说明，如何在“Z＋Z 超级画板”的程序工作区计算统计量．

最直接的办法是把数据和公式直接输入．如 12.1.3 的例 3，要计算 10 个数据

9.5，9.9，9.9，9.9，9.8，9.7，9.5，9.3，9.6，9.6

的均值并用 mx 表示，就在程序工作区键入：

mx＝(9.5＋9.9＋9.9＋9.9＋9.8＋9.7＋9.5＋9.3＋9.6＋9.6)/10;

按 Ctrl＋Enter 键执行（下同）得到

≫(967)/(100) #

如果想要用小数表示，可执行

Float(1);

返回：

≫计算结果显示浮点数 #

再键入

mx;

执行得

≫(967)/(100)=9.67 #

要计算这组数据的方差 s2，可键入

s2=(9.5^2+9.9^2+9.9^2+9.9^2+9.8^2+9.7^2+9.5^2+9.3^2+9.6^2+9.6^2)/10-mx^2;

执行得到

≫(381)/(10000)=0.0381 #

计算标准差则再键入：

sx=s2^0.5;

执行即得到

≫((381)^(1/2))/(100)=0.195192 #

用这种方法计算的好处，在于能在计算过程中多次复习这些统计量的意义和公式．用了计算机，计算统计量的主要工作就是输入数据．为了避免多次输入“+”号和平方运算符号“^2”的重复性工作，可以做一个“模板”，也就是先打一行符号，如

(^2+^2+^2+^2+^2+^2+^2+^2+^2+^2+^2+^2+^2+^2+^2+^2)

复制一下，用时随时粘贴，再在符号“^”前面依次键入数据就是了．

小结与复习

总体和个体

1. 样本：是从总体中抽取的一部分个体，也称为观测数据.

2. 总体均值：总体的平均，常用 μ 表示.

3. 样本均值：样本的平均，常用 $\bar{x}$ 表示. 样本平均是总体均值的估计.

4. 均值的性质：

(1) 如果每个数据增加 b，均值也增加 b；

(2) 如果每个数据扩大 a 倍，均值也扩大 a 倍.

5. 总体方差：当 y_1，y_2，…，y_N 是总体的全部个体，μ 是总体均值时，总体方差

$$\sigma^2=\frac{(y_1-\mu)^2+(y_2-\mu)^2+\cdots+(y_N-\mu)^2}{N}.$$

6. 样本方差：用 $\bar{x}$ 表示样本 x_1，x_2，…，x_n 的均值时，样本方差

$$s^2=\frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\cdots+(x_n-\bar{x}^2)}{n}.$$

样本方差是总体方差的估计.

7. 方差的性质：

(1) 方差描述数据向均值的集中程度：方差越小，个体向均值集中得越好. 方差也描述数据的整齐程度或波动幅度：方差越小，数据就越整齐.

(2) $s^2=\frac{1}{n}(x_1^2+x_2^2+\cdots+x_n^2)-\bar{x}^2$.

8. 标准差：标准差是方差的算术平方根. 数据带有单位时，标准差的单位和数据的单位是一致的.

抽样调查方法

1．抽样调查的必要性：在很多实际问题中，采用抽样的方法来确定总体性质不仅是必要的，也是必须的．

2．随机抽样：指总体中的每个个体都有相同的机会被抽中．

3．简单随机抽样：无放回随机抽样．

4．随机抽样的重要性：不进行随机抽样，就可能得到有偏的样本．利用有偏的样本很容易得到错误的结论．

5．随机数：是从 1 至 n 个号码中用简单抽样方法抽到的 m 个号码．随机数可以在计算机上产生．

6．分层抽样：将总体分成若干层（子总体），然后在每层中独立地进行简单随机抽样．

7．分层抽样的特点：

（1）分层抽样在获得总体均值估计的同时，也得到各层的均值估计；

（2）将差别不大的个体分在同一层，使得分层抽样得到的样本更具有代表性，从而提高估计的准确度；

（3）抽样调查的实施更加方便，调查数据的收集、处理也更加方便．

8．系统抽样：当总体中的个体按一定的方式排列，在规定的范围内随机抽取一个个体，然后按照制定好的规则确定其他个体的抽样方法．

9．系统抽样的特点：实施简单，如果了解总体中个体排列的规律，设计合适的系统抽样规则可以增加估计的精度．

用样本分布估计总体分布

1．频率分布表，频率分布直方图，频率折线图和数据的茎叶图都是用来展示样本的分布性质的，这些分布性质是总体分布性质的估计或近似．

2．绘制频率分布表时，数据分段个数 K 可以参考经验公式

$$K=1+4\lg n$$

确定．但是经验公式只起参考作用．实际应用时，应当根据样本量的大小和数据的特点以及分析的要求灵活确定．

说明：由于频率分布表的制作没有统一的数据分段方法，所以对相同的数据，可以作出不同的频率分布表，因而也可以作出不同的直方图和频率折线图．但是好的频率分布表、直方图和频率折线图应当是简单明了的．

数据的相关性

样本量是 n 的成对观测数据是用

$$(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$$

表示的．这里，对固定的 i，x_i 和 y_i 或是来自相同的个体，或是同一次试验的观测数据．对 $i\neq j$，(x_i, y_i) 和 (x_j, y_j) 或是来自不同的个体，或是不同次试验的观测数据．

1．当 (x_1, y_1)，(x_2, y_2)，…，(x_n, y_n) 十分明显地集中在一条上升的直线附近时，称 x 和 y 是高度正相关的，当它们分布在一条上升的直线附近时称 x 和 y 是中度正相关的．高度正相关和中度正相关统一称为正相关．

2．当 (x_1, y_1)，(x_2, y_2)，…，(x_n, y_n) 十分明显地集中在一条减少的直线附近时，称 x，y 是高度负相关的；当它们分布在一条减少的直线附近时，称 x 和 y 是中度负相关的．高度负相关和中度负相关统一称为负相关．

3．当 x 和 y 正相关或负相关时，称 x 和 y 是相关的．

4．回归直线：当 $\{x_i\}$ 和 $\{y_i\}$ 有相关性时，可以用一条直线描述数据 $\{x_i\}$ 和 $\{y_i\}$ 的关系，这条直线就是回归直线．用

$$l:\ y=bx+a$$

表示这条回归直线时，其中的 a，b 可以用下面的公式进行计算：

$$b=\frac{s_{xy}}{s_x^2},\quad a=\bar{y}-b\bar{x}.$$

5．用回归直线进行预测：得到了回归直线后，只要 $\{x_i\}$ 与

$\{y_i\}$ 高度相关，对于新的 x，就可以用回归直线上的点 $y=bx+a$ 作为 y 的预测值.

复习题十二

学而时习之

1. 在调查某城市的居民对自来水调价的意见时，能否只在该市的娱乐场所进行随机抽样？原因是什么？能否只在该市的公共汽车站进行随机抽样？原因是什么？

2. 以下是对某个公交车站候车的乘客候车时间的调查数据（单位：min）.

21	1	22	19	9	23	15	19	5	2	7	13
4	20	24	18	6	16	19	5	1	21	16	2
14	16	7	2	5	3	4	19	5	14	13	20
18	9	7	22	10	9	9	23	11	1	18	21

计算样本均值和样本标准差.

3. 把 2，3，4，5，9 的茎视为 0，为以下两组数据制作双茎叶图.

数据 1：

2	3	3	4	5	5	6	8	9	10	10	11
11	13	13	14	15	15	15	18	19	19	19	20
22	23	25	26	27	27	28	30	30	32	34	34
35	36	37	37	39	40	41	44	44	47	49	49

数据 2：

4	4	8	8	9	11	11	13	15	16	17	17
17	17	18	24	25	25	28	30	31	31	31	35
39	42	48	50	53	55	56	57	59	60	62	62
65	68	70	75	75	76	76	76	77	77	78	89

4. 要在全年级 450 名同学中随机选取 45 人参加暑假的夏令营时，完成以下工作：

(1) 设计一个随机抽样方案；

（2）设计一个系统抽样方案；

（3）设计一个分层抽样方案，使得选取出男生 23 名，女生 22 名；

（4）如果全年级有 9 个班，设计一个分层抽样方案，使得各班随机选取 5 人.

温故而知新

5. 从某个渔场通过简单随机抽样方法检查了 30 条鱼的质量（单位：g）如下：

254	298	320	325	347	354	362	367	369	372
379	380	382	386	391	397	397	411	412	419
422	435	438	443	447	452	480	506	524	539

（1）制作频率分布表；

（2）制作频率直方图；

（3）制作频率折线图；

（4）制作茎叶图；

（5）计算样本均值；

（6）计算样本方差.

6. 按以下要求分别设计随机抽样调查方案，调查本年级的近视率.

（1）要求随机调查 50 个同学；

（2）要求调查男女同学各 25 人；

（3）要求调查每班 10 个同学；

（4）要求调查每班的男女生各 5 人.

7. 2000 年的 3—10 月北京和广州的月平均气温（单位：℃）记录如下：

月份	3	4	5	6	7	8	9	10
北京 x	8	15	20	27	30	26	22	13
广州 y	19	23	26	28	29	28	27	25

为北京和广州月平均气温建立回归直线.

上下而求索

8. 下面是世界上 10 个地区的人均收入和受教育比例的数据：

人均收入 x	61	165	125	645	398	208	289	311	246	86
受教育比例 y	6	43	50	87	80	71	30	77	96	77

（1）为数据建立回归直线；

（2）绘制数据的散点图和回归直线；

（3）对 $x=60$，70，200 分别预测 y 的值；

（4）对于 $y=85$ 预测 x 的值.

9. 下面是某种合成纤维的拉伸强度 y 和拉伸倍数 x 之间的测量数据.

x	1.9	2.0	2.1	2.5	2.7	2.7	3.5	3.5	4.0	4.0	4.5	4.6
y	1.4	1.3	1.8	2.5	2.8	2.5	3.0	2.7	4.0	3.5	4.2	3.5

（1）为数据建立回归直线；

（2）绘制数据的散点图和回归直线；

（3）对 $x=5.0$，5.2 分别预测 y 的值；

（4）对 $y=4.5$ 预测 x 的值.

第 13 章

概 率

沙场百胜古来稀，
九密一疏已足奇.
祸福偶然存概率，
风云变幻识玄机.

在考虑一个未来事件是否会发生的时候，人们常关心该事件发生的可能性的大小. 概率就是用来衡量一个未来事件发生的可能性大小的度量. 概率通过对简单随机事件的研究，将人们逐步引入复杂随机现象规律的研究. 概率是研究复杂随机现象规律的有效方法和工具.

投掷一枚均匀的硬币，出现正面朝上的机会是 50 %. 在多次投掷这枚硬币后，正面朝上和反面朝上出现的次数大致相同. 南非数学家凯瑞（Kerrich）是在非常困难的情况下发现这一点的. 他的掷币试验是在德国人的集中营里进行的. 二战爆发时他访问哥本哈根，德国人入侵丹麦时他被关进集中营，在那里度过了漫长的岁月. 为了消磨时间，他一次次地掷硬币，并记录了下面的结果：

掷币次数 N	正面次数 n	频率 $f=n/N$	掷币次数 N	正面次数 n	频率 $f=n/N$
10	4	0.400	600	312	0.520
20	10	0.500	700	368	0.526
30	17	0.567	800	413	0.516
40	21	0.525	900	458	0.509
50	25	0.500	1 000	502	0.502
60	29	0.483	2 000	1 013	0.507
70	32	0.457	3 000	1 510	0.503
80	35	0.438	4 000	2 029	0.507
90	40	0.444	5 000	2 533	0.507
100	44	0.440	6 000	3 009	0.502
200	98	0.490	7 000	3 516	0.502
300	146	0.487	8 000	4 034	0.504
400	199	0.498	9 000	4 538	0.504
500	255	0.510	10 000	5 067	0.507

从上述掷币结果看出，随着掷币次数的增加，正面朝上的频率稳定在 0.5 附近. 我们将这个性质称为**频率的稳定性**.

从学习过的概率知识知道，投掷一枚硬币，出现正面朝上的概率是 0.5. 在多次重复试验中，正面朝上的频率稳定在正面朝上的概率附近.

13.1　试验与事件

13.1.1　事　件

投掷一枚硬币，用 H 表示硬币正面朝上，用 T 表示硬币反面朝上，则试验有两个可能的结果：H 和 T. 我们把 H 和 T 叫作试验的元素，把集合 $\{H,T\}$ 叫作试验的全集.

投掷一枚骰子，用 1 表示掷出点数 1，用 2 表示掷出点数 2，…，用 6 表示掷出点数 6，则试验的可能结果是 1，2，3，4，5，6. 我们称这 6 个数是试验的元素. 称

$$\{j \mid j=1,2,\cdots,6\}$$

为试验的全集.

Ω　音 omega.

这里 ω(音：omega) 是 Ω 的小写.

对于一个试验，我们将该试验的可能结果称为元素，称元素构成的集合为试验的全集.

以后总用 Ω 表示试验的全集，用 ω 表示 Ω 的元素.

例 1　同时投掷一枚 1 角的硬币和一枚 1 元的硬币，写出试验的元素和全集.

解　试验一共有 4 个元素，它们是

HH：1 角硬币正面朝上，1 元硬币正面朝上；

HT：1 角硬币正面朝上，1 元硬币反面朝上；

TH：1 角硬币反面朝上，1 元硬币正面朝上；

TT：1 角硬币反面朝上，1 元硬币反面朝上.

全集是 $\Omega=\{HH,HT,TH,TT\}$.

在上面的例子中，HT 和 TH 是不同的元素.

例 2　投掷两枚硬币，写出试验的元素和全集.

解　将一个硬币视为硬币 1，另一个硬币视为硬币 2. 试验有 4 个元素，它们是

HH：硬币 1 正面朝上，硬币 2 正面朝上；

HT：硬币 1 正面朝上，硬币 2 反面朝上；

TH：硬币 1 反面朝上，硬币 2 正面朝上；

TT：硬币 1 反面朝上，硬币 2 反面朝上.

全集是 $\Omega=\{HH,HT,TH,TT\}$.

和例 1 中的情况相同，这里 HT 和 TH 也是不同的元素.

投掷一枚骰子的全集是

$$\Omega=\{j \mid j=1,2,\cdots,6\}.$$

用 $A=\{3\}$ 表示掷出 3 点，A 是 Ω 的子集. 我们称 A 是**随机事件**(random event)，简称为事件. 掷出 3 点，就称事件 A 发生，否则称事件 A 不发生.

用 $B=\{2,4,6\}$ 表示掷出偶数点，B 是 Ω 的子集，我们也称 B 是随机事件，简称为事件. 当掷出偶数点，称事件 B 发生，否则称事件 B 不发生. 事件 B 发生和掷出偶数点是等价的.

当 Ω 是试验的全集，我们称 Ω 的子集 A 是 Ω 的事件，简称为事件 (event). 当试验的元素（即试验结果）ω 属于 A，就称事件 A 发生，否则称 A 不发生.

对于全集 Ω，A 是事件和 $A\subseteq\Omega$ 等价. 元素 $\omega\in A$ 和事件 A 发生等价.

空集$\varnothing$也是 Ω 的子集，所以空集$\varnothing$是事件. 空集$\varnothing$中没有元素，永远不会发生，所以我们称$\varnothing$是**不可能事件**.

Ω 也是 Ω 的子集，并且包括了所有的元素，所以必然发生. 我们称全集 Ω 是**必然事件**.

事件是随机事件的简称.

元素作为试验的可能结果，又被统计学家们称为试验的**样本点**(sample outcome) 或基本事件.

试验的全集 Ω 又被统计学家们称为**样本空间**(sample space).

练 习

1. 叙述什么是试验的元素和全集.
2. 简述事件和全集的关系.
3. 简述事件和元素的关系.

习题 1

学而时习之

1. 口袋中有标号 1～3 的球各 1 个. 为以下的试验写出全集.

(1) 从中任取 1 个;

(2) 从中一次随机地取出 2 个.

2. 投掷一枚骰子和一枚硬币，写出全集.

3. 同时投掷一枚骰子和一枚硬币，写出以下事件.

(1) 硬币是正面，骰子的点数是奇数;

(2) 硬币是正面，骰子的点数是偶数;

(3) 硬币是正面;

(4) 骰子的点数是 5.

13.1.2　事件的运算

由于事件就是集合，所以对事件可以进行并、交和补的运算.

例 1　投掷两枚骰子，一枚是红色，一枚是蓝色. 写出全集和以下事件.

(1) $A=$ “红骰子的点数是 2”;

(2) $B=$ “蓝骰子的点数是 3”;

(3) $A\cap B$;

(4) $A\cup B$.

解　用(i,j)表示红色骰子的点数是 i，蓝色的点数是 j. 试验的全集是

$$\Omega=\{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),$$
$$(2,1),(2,2),(2,3),(2,4),(2,5),(2,6),$$

(3，1)，(3，2)，(3，3)，(3，4)，(3，5)，(3，6)，

(4，1)，(4，2)，(4，3)，(4，4)，(4，5)，(4，6)，

(5，1)，(5，2)，(5，3)，(5，4)，(5，5)，(5，6)，

(6，1)，(6，2)，(6，3)，(6，4)，(6，5)，(6，6)}.

根据事件的定义，得到

(1) $A=\{(2,1),(2,2),(2,3),(2,4),(2,5),(2,6)\}$；

(2) $B=\{(1,3),(2,3),(3,3),(4,3),(5,3),(6,3)\}$；

(3) $A\cap B=\{2,3\}$ ＝“红骰子是 2 点，蓝骰子是 3 点”；

(4) $A\cup B=\{(2,1),(2,2),(2,3),(2,4),(2,5),(2,6),(1,3),(3,3),(4,3),(5,3),(6,3)\}$

＝“红骰子是 2 点或蓝骰子是 3 点”.

用 $\Omega\setminus A$ 表示 A 的补集.

在例 1 中，$\Omega\setminus A$ 表示红骰子的点数不是 2.

对于试验的全集 Ω 和事件 A，由于 A 和 $\Omega\setminus A$ 有且只能有一个发生，所以我们称 $\Omega\setminus A$ 是 A 的**对立事件**.

例 2　铅笔盒中有圆珠笔 3 支，钢笔 2 支. 从中无放回地任取 3 支，用集合 A，B，C 表示下面 (1)，(2)，(3) 中的事件.

(1) 3 支都是圆珠笔；

(2) 恰有 2 支圆珠笔；

(3) 恰有 1 支圆珠笔；

(4) 用 A，B，C 表示 Ω；

(5) 解释事件 $A\cup B$，$A\cap B$，$A\setminus B$，$\Omega\setminus A$ 的含义.

解　将 3 支圆珠笔标号 1，2，3，将 2 支钢笔编号 1，2. 用 y_i 和 g_j 分别表示取出的有第 i 支圆珠笔和第 j 支钢笔. 用 $y_1y_2y_3$ 表示取出的是 1 号，2 号和 3 号圆珠笔，$y_1y_2g_1$ 表示取出的是 1 号、2 号圆珠笔和 1 号钢笔，…按照事件的定义，得到

(1) $A=\{y_1y_2y_3\}$.

(2) $B=\{y_1y_2g_1, y_1y_2g_2, y_1y_3g_1, y_1y_3g_2, y_2y_3g_1, y_2y_3g_2\}$.

(3) $C=\{y_1g_1g_2, y_2g_1g_2, y_3g_1g_2\}$.

(4) 因为必有事件 A,B,C 之一发生，所以全集 $\Omega=A\cup B\cup C$.

（5）$A\cup B=$ “至少有 2 支圆珠笔”；

$A\cap B=\varnothing=$ “不可能事件”；

$A\setminus B=A=$ “3 支都是圆珠笔”；

$\Omega\setminus A=$ “至少有 1 支钢笔”.

在例 2 中，事件 A，B 的交集是空集，所以 A 发生，B 就不能发生；B 发生，A 就不能发生.

当 $A\cap B=\varnothing$，我们称 A，B **互斥**.

练 习

投掷 3 枚硬币，观察结果. 写出全集，分别用集合 A，B，C 表示以下（1），（2），（3）中的事件.

（1）恰好两个正面；

（2）至少一个正面；

（3）都是反面；

（4）计算 $B\setminus A$，$B\setminus C$，$\Omega\setminus A$，并解释它们的含义.

习题 2

学而时习之

1. 投掷 4 枚硬币，观察结果. 不写出全集，直接用集合 A，B，C 表示以下（1），（2），（3）中的事件.

（1）至少 1 个反面朝上；

（2）至少 2 个反面朝上；

（3）恰好 2 个反面朝上；

（4）计算 A，$A\cap C$，$\Omega\setminus A$，并解释含义.

2. 盒中有标号 1～3 的白球各 1 个，标号 1～2 的黑球各 1 个. 从中倒出 3 个，观

察结果. 写出全集，用集合 A,B,C 表示下面(1),(2),(3)中的事件.

(1) 3 个都是白球；

(2) 至少 2 个白球；

(3) 至少 1 个白球；

(4) 计算 $A\cup B$，$A\cap B$，$A\setminus B$，$C\setminus B$，并解释它们的含义.

13.2 概率及其计算

13.2.1 古典概率模型

1. 概率的定义.

投掷一枚均匀的硬币，全集 $\Omega=\{H,T\}$ 中有两个元素. 事件 $A=\{H\},B=\{T\}$各有一个元素. 根据已有的概率知识，

$$P(A)=\frac{1}{2}=\frac{A\text{ 中元素数}}{\Omega\text{ 中元素数}},$$

$$P(B)=\frac{1}{2}=\frac{B\text{ 中元素数}}{\Omega\text{ 中元素数}}.$$

投掷一枚均匀的骰子，全集是

$$\Omega=\{j\mid j=1,2,\cdots,6\}.$$

$A=\{j\}$表示掷出的点数是 j，$B=\{2,4,6\}$表示掷出偶数点. 根据已有的概率知识，有

$$P(A)=\frac{A\text{ 中元素数}}{\Omega\text{ 中元素数}}=\frac{1}{6},\qquad P(B)=\frac{B\text{ 中元素数}}{\Omega\text{ 中元素数}}=\frac{1}{2}.$$

从以上的例子引出概率的如下定义.

定义 设试验的全集 Ω 有 n 个元素，且每个元素发生的可能性相同. 当 Ω 的事件 A 包含了 m 个元素时，称

$$P(A)=\frac{m}{n}$$

为事件 A 发生的概率，简称为 A 的**概率** (probability).

我们把上述定义描述的概率模型称为古典概率模型，简称为**古典概型**.

注 以下所述的硬币、骰子等都是均匀的.

例 1 投掷 3 枚硬币. 计算以下事件的概率.

(1) 至少 1 个反面朝上；

(2) 至少 2 个反面朝上；

(3) 恰好 2 个反面朝上.

解 试验的全集

$\Omega=\{HHH,HHT,HTH,THH,HTT,THT,TTH,TTT\}$.

(1) 用 A 表示至少有一个反面，则 A 中的元素至少含一个 T.

$$A=\{HHT,HTH,HTT,THH,THT,TTH,TTT\}$$

含有 7 个元素，于是 $P(A)=7/8$.

(2) 用 B 表示至少两个反面，则

$$B=\{HTT,TTH,THT,TTT\}$$

含有 4 个元素，于是 $P(B)=4/8=1/2$.

(3) 用 C 表示恰好两个反面朝上，则

$$C=\{HTT,TTH,THT\}$$

含有 3 个元素，于是 $P(C)=3/8$.

事件作为集合经过并、交、差和补的运算后得到的结果还是事件，于是可以计算经过集合运算后的事件的概率.

例 2 在例 1 中，计算事件 $A\cup B$，$A\cap C$，$B\setminus C$，$\Omega\setminus C$ 发生的概率.

解 全集 Ω 有 8 个元素.

$A\cup B=A$ 含有 7 个元素，

$A\cap C=C$ 含有 3 个元素，

$B\setminus C=\{TTT\}$含有 1 个元素，

$\Omega\setminus C=\{HHH,HHT,HTH,THH,TTT\}$含有 5 个元素. 所以有

$$P(A\cup B)=\frac{7}{8},\ P(A\cap C)=\frac{3}{8},\ P(B\setminus C)=\frac{1}{8},\ P(\Omega\setminus C)=\frac{5}{8}.$$

2. 概率的性质.

概率有如下的简单性质：

(1) $0\leqslant P(A)\leqslant 1$（概率总是 $[0, 1]$ 中的数）；

(2) $P(\Omega)=1$（必然事件的概率是 1）；

(3) $P(\varnothing)=0$（不可能事件的概率是零）.

在一副扑克的 54 张中随机抽取 1 张. 用 A 表示得到的是草花，用 B 表示得到的是黑桃，则 A，B 互斥. 全集 Ω 有 54 个元素，A 和 B 分别有 13 个元素. 不用写出具体的全集 Ω 和事件 A，B，也可以直接计算出

草花就是我们常说的梅花.

$$P(A)=\frac{13}{54},\quad P(B)=\frac{13}{54}.$$

由于 $A\cup B$ 也是事件，含有 26 个元素，所以

$$P(A\cup B)=\frac{26}{54}=\frac{13}{54}+\frac{13}{54}=P(A)+P(B).$$

概率的加法公式： 如果 Ω 的事件 A，B 互斥，则

$$P(A\cup B)=P(A)+P(B).$$

我们把概率的加法公式称为概率的**可加性**. 可加的前题是两个事件互斥.

证 设 Ω 有 n 个元素，A 有 m 个元素，B 有 k 个元素. 则 $P(A)=m/n$，$P(B)=k/n$. 由于 A，B 互斥，所以

$$A\cup B \text{ 中元素个数}=A \text{ 中元素个数}+B \text{ 中元素个数}=m+k.$$

于是得到

$$\begin{aligned}P(A\cup B)&=\frac{m+k}{n}=\frac{m}{n}+\frac{k}{n}\\&=P(A)+P(B).\end{aligned}$$

对立事件的概率公式： 如果 A 是全集 Ω 的事件，则

$$P(\Omega\backslash A)=1-P(A).$$

证 $\Omega\backslash A$ 和 A 互斥，并且 $\Omega=(\Omega\backslash A)\cup A$. 由概率的加法公式得到

$$P(\Omega)=P(\Omega\backslash A)+P(A).$$

再利用 $P(\Omega)=1$ 得到 $1=P(\Omega\backslash A)+P(A)$，于是 $P(\Omega\backslash A)=1-P(A)$.

例 3　在一副扑克的 54 张中随机抽取 1 张.

(1) 计算抽到是草花或黑桃的概率;

(2) 计算抽到的不是草花的概率;

(3) 计算抽到的不是草花也不是黑桃的概率.

解　用 A 表示抽到的是草花，用 B 表示抽到的是黑桃，则 $A\cup B$ 表示抽到的是草花或黑桃，并且 A，B 互斥. $P(A)=13/54$,$P(B)=13/54$,所以

$$P(A\cup B)=P(A)+P(B)=\frac{13}{54}+\frac{13}{54}=\frac{13}{27}.$$

$\Omega\backslash A$ 表示抽到的不是草花,是 A 的对立事件,所以

$$P(\Omega\backslash A)=1-P(A)=1-\frac{13}{54}=\frac{41}{54}.$$

$C=A\cup B$ 表示抽到的是草花或黑桃，$\Omega\backslash C$ 表示抽到的不是草花也不是黑桃.

$$P(\Omega\backslash C)=1-P(C)=1-P(A\cup B)=1-\frac{13}{27}=\frac{14}{27}.$$

例 4　袋中有红球和白球各 1 个，每次抽 1 个，有放回地随机抽取 3 次. 计算:

(1) $A=$ “至少有 1 个红球”的概率;

(2) $B=$ “至少有 1 个白球”的概率;

(3) 有红球或有白球的概率.

解　用 BHB 表示第 1、第 2 和第 3 次分别取到白、红、白球等. 全集是

$$\Omega=\{BBB,BBH,BHB,HBB,BHH,HHB,HBH,HHH\},$$

$$A=\{BBH,BHB,HBB,BHH,HHB,HBH,HHH\},$$

$$P(A)=\frac{7}{8}=0.875.$$

由于红球和白球处于对称的地位,所以 $P(B)=P(A)=0.875$. $A\cup B$ 表示有红球或有白球,这是必然事件,所以 $P(A\cup B)=1$.

例 5　彩票的中奖率是 1/2，每次抽 1 张，有放回地随机抽取 3 次. 计算:

（1）$A=$“至少抽中 1 次”的概率；

（2）1 次也没抽中的概率.

解 将带奖的彩票视为红球，不带奖的彩票视为白球，则 A 等价于至少抽中 1 个红球. 由例 3 的结论知道 $P(A)=0.875$. $\Omega\backslash A$ 是 1 次也没抽中的概率，$P(\Omega\backslash A)=1-0.875=0.125$.

例 6 有 10 万张彩票，中奖率是 1/2. 每次抽 1 张，无放回地随机抽取 3 次，计算至少抽中 1 次的概率.

解 由于彩票数量很大，抽取一两张基本不会影响彩票的中奖比例，所以无放回抽奖的中奖概率和有放回抽奖的中奖概率基本是一样的. 要求的概率仍然是 0.875.

有人认为既然每次抽中的概率是 1/2，抽两次必然抽中，这是不对的. 上面的例 3、例 4 告诉我们，抽奖 3 次时，至少抽中 1 次的概率只有 0.875.

完全相同的道理，当彩票的中奖率是 1/100 时，你购买 100 张彩票中奖的概率是严格小于 1 的（实际上只有 0.634）.

练 习

投掷两枚骰子，一枚是红色，一枚是蓝色. 计算以下事件的概率.

（1）$A=$“两枚骰子的点数相同”；

（2）$B=$“红色骰子的点数小于蓝色骰子的点数”；

（3）$C=$“两枚骰子的点数之和是 6”；

（4）$A\cup B$，$A\cup C$，$C\backslash B$.

习题 3

学而时习之

1. 假设每个人的生日在一年的 365 天中是等可能的. 在全校随机挑选一名同学，计算以下事件的概率.

 (1) 该同学的生日在 5 月份；

 (2) 该同学的生日在 5 月或 7 月份；

 (3) 该同学的生日是 1 日；

 (4) 该同学的生日是 1 日或 2 日.

2. 一批产品有 100 个，其中含有 10 个次品，从中随机抽取一个. 计算：

 (1) 这件产品是次品的概率；

 (2) 这件产品是正品的概率；

 (3) 这件产品是次品或是正品的概率.

3. 某电视台要招聘两名播音员，现在有三位符合条件的女士和两位符合条件的男士前来应聘. 如果每位应聘人员被录用的概率相同，计算以下概率.

 (1) 一位男士和一位女士被录用的概率；

 (2) 两位男士被录用的概率；

 (3) 两位女士被录用的概率.

4. 投掷两枚骰子，不用写出全集，计算以下概率.

 (1) 两枚骰子的点数相同；

 (2) 两枚骰子的点数之和是 6；

 (3) 两枚骰子的点数之和不是 6；

 (4) 至少一枚骰子的点数是 3.

5. 1 万张彩票中有一个大奖.

 (1) 从中抽取 1 张时，中奖的概率是多少？不中奖的概率是多少？

 (2) 从中有放回地随机抽取 2 张，中奖的概率是多少？不中奖的概率是多少？

 (3) 从中无放回地随机抽取 2 张时，中奖的概率是多少？不中奖的概率是多少？

6. 如果全集 Ω 的事件 A，B，C 两两互斥，用 $A\cup B\cup C$ 表示事件 A，B，C 中至

少有一个发生．证明：

$$P(A\cup B\cup C)=P(A)+P(B)+P(C).$$

7. 彩票的中奖率是 1/3，每次抽 1 张，有放回地随机抽取 3 张．计算至少抽中 1 张的概率．

13.2.2　几何概率

例 1　在区间 $[0,3)$ 中随机投掷一个质点，分别求质点落在区间 $[0,1)$ 和 $[1,2)$ 中的概率．

解　用 $A=[0,1)$ 表示质点落在 $[0,1)$ 中，用 $B=[1,2)$ 表示质点落在 $[1,2)$ 中，用 $C=[2,3)$ 表示质点落在 $[2,3)$ 中．把 A,B,C 看作试验的元素时，试验的全集 $\Omega=A\cup B\cup C=[0,3)$．$A,B,C$ 发生的可能性相同，所以

$$P(A)=\frac{1}{3}=\frac{A\text{ 的长度}}{\Omega\text{ 的长度}}.$$

$$P(B)=\frac{1}{3}=\frac{B\text{ 的长度}}{\Omega\text{ 的长度}}.$$

在概率的语言中，我们将“随机”解释成“等可能”．

> **几何概率定义 1**　设试验的全集 Ω 是长度为正数的区间，A 是 Ω 的子区间．如果试验的结果随机地落在 Ω 中，则称
>
> $$P(A)=\frac{A\text{ 的长度}}{\Omega\text{ 的长度}}$$
>
> 为事件 A 发生的概率，简称为 A 的概率．

在几何概率定义 1 中，并不指定所述的 Ω 和 A 是开区间、闭区间，还是半开半闭的区间．这是因为区间 $(a,\ b)$，$[a,\ b)$，$(a,\ b]$，$[a,\ b]$ 有相同的长度．

例 2　公共汽车在 0～5 min 内随机地到达车站．

(1) 求汽车第 3 min 到达车站的概率；

(2) 求汽车在 1～3 min 之间到达的概率．

解 试验的全集是 $\Omega=[0, 5]$，集合 $A=[3, 3]$ 表示汽车在第 3 min时到达，$B=[1, 3]$ 表示汽车在 1～3 min 之间到达．根据几何概率定义：

$$P(A)=\frac{0}{5}=0, \quad P(B)=\frac{2}{5}.$$

在例 2 中也可以用 $\Omega=(0, 5)$ 表示全集，用 $B=(1, 3)$ 表示汽车在 1～3 min 之间到达．计算的 $P(B)$是一样．

下面把几何概率的定义推广到平面上．我们把平面上的矩形、圆、椭圆等统一称为**区域**．

> **几何概率定义 2** 设试验的全集 Ω 是面积为正数的区域，A 是 Ω 的子区域，如果试验的结果随机地落在 Ω 中，则称
>
> $$P(A)=\frac{A \text{ 的面积}}{\Omega \text{ 的面积}}$$
>
> 为事件 A 发生的概率，简称为 A 的概率．

几何概率也有如下的基本性质．

(1) $0 \leqslant P(A) \leqslant 1$ （概率总是$[0,1]$中的数），

(2) $P(\Omega)=1$ （必然事件的概率是 1），

(3) $P(\varnothing)=0$ （不可能事件的概率是零），

(4) 如果 A,B 互斥，则 $P(A \cup B)=P(A)+P(B)$，

(5) $P(A)+P(\Omega \backslash A)=1$ （对立事件概率之和等于 1）．

例 3 设雨点等可能地落在半径是 1 m 的圆 Ω 中，A 是半径为 0.5 m 的圆（如图 13－1）．

(1) 计算雨点落在小圆内的概率；

(2) 计算雨点落在小圆外的概率．

解 大圆的面积是 π m²，雨点等可能地落入大圆．小圆的面积是 $\pi\, 0.5^2$ m²，雨点落入小圆的概率是

$$P(A)=\frac{A \text{ 的面积}}{\Omega \text{ 的面积}}=\frac{\pi 0.5^2}{\pi}=0.25.$$

根据几何概率的性质 5，雨点落入小圆外的概率是

$P(\Omega\backslash A)=1-0.25=0.75.$

例 4　陨石等可能地溅落在方圆为 200 km^2 的区域内，该区域内有面积为 80 km^2 的湖泊. 求陨石溅落在湖泊中的概率.

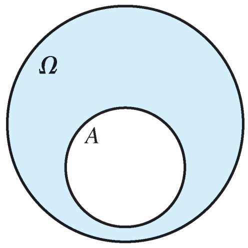


图 13－1

解　全集 Ω 的面积是 200 km^2，湖泊 A 的面积是 80 km^2. 元素等可能地落在 Ω 中，由几何概率的定义得到

$$P(A)=\frac{A\text{ 的面积}}{\Omega\text{ 的面积}}=\frac{80}{200}=0.4.$$

陨石溅落在湖泊中的概率是 0.4.

练　习

每天的整点（如 9 时，10 时，11 时等）北京站都有列车发往天津. 一位乘客在 9 时至 10 时之间随机到达北京站. 计算：

（1）他候车多于 20 min 的概率；

（2）他候车恰好 15 min 的概率；

（3）他候车少于 30 min 的概率；

（4）他候车时间在 25～45 min 之间的概率.

习题 4

温故而知新

1. 一只麻雀随机地落在面积是 400 m^2 的广场上觅食. 广场内有一个长 20 m、宽 5 m的草坪，还有一个半径是 5 m 的花坛. 计算：

（1）麻雀落在草坪中的概率；

（2）麻雀落在花坛内的概率；

(3) 麻雀落在草坪或花坛内的概率；

(4) 麻雀落在草坪或花坛外的概率.

2. 在长方形 $\Omega=A\cup B\cup C\cup D$ 中随机投掷一个质点（见图 13－2）. 计算：

(1) $P(A)$；

(2) $P(A\cup B)$；

(3) $P(A\cup C\cup D)$；

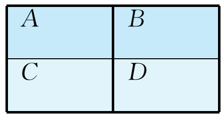


图 13－2

(4) 证明 $P((A\cup B)\cap(A\cup C))=P(A\cup B)P(A\cup C)$.

3. 在区间 [0，1] 中随机地投掷一点，计算该点落在 [0，0.3] 中的概率.

4. 在区间 [0，1] 中随机地取两点，用 A 表示它们的平方和小于 1.

(1) 写出试验的全集 Ω；

(2) 用集合表示出事件 A；

(3) 计算 $P(A)$.

5. 在区间[1,2]中随机地投掷两个点,用 B 表示它们的差的绝对值小于 1/3.

(1) 写出试验的全集 Ω；

(2) 用集合表示出事件 B；

(3) 计算 $P(B)$.

6. 两人在某天的 1 时至 2 时间各自独立随机到达某地会面，先到者等候 20 min 后离去.

(1) 写出试验的全集 Ω；

(2) 用集合表示出事件 $B=$ “两人相遇”；

(3) 计算这两人能相遇的概率.

13.3 频率与概率

设 Ω 是某个试验的全集，A 是 Ω 的事件. 在相同的条件下将该试验独立地重复 N 次，我们称

$$f_N=\frac{N\text{ 次试验中 }A\text{ 发生的次数}}{N}$$

是 N 次独立重复试验中，事件 A 发生的**频率**.

这个结论首先由伯努利给出数学的证明.

理论和事实都证明：在相同的条件下，将一试验独立重复 N 次，

用 f_N 表示事件 A 在这 N 次试验中发生的频率. 当 N 增加时，f_N 将在一个固定的数值 p 附近波动，这个数值 p 就是事件 A 的概率 $P(A)$. 于是，f_N 是 $P(A)$的估计.

历史上许多著名的统计学家对概率和频率的关系进行过验证. 他们的试验结果总结在表 13.1 中.

表 13.1

试验者	掷币次数 N	正面朝上次数	频率 f_N
德·摩根	2 048	1 061	0.518 1
蒲丰	4 040	2 048	0.506 9
凯瑞	7 000	3 516	0.502 2
凯瑞	9 000	4 538	0.504 2
费勒	10 000	4 979	0.497 9
皮尔逊	12 000	6 019	0.501 6
皮尔逊	24 000	12 012	0.500 5
罗曼诺夫斯基	80 640	40 173	0.498 2

现在的随机试验工作可以在计算机上方便地进行.

例 1　表 13.2 是用计算机进行的投掷一枚均匀的骰子的试验总结. 其中 N 是试验的次数，表中的百分数是频率. 例如表中第 3 行第 2 列的 15.00 %，表示试验次数 $N=10^2$ 时，点数 1 出现的频率是 15.00 %.

表 13.2

点数	$N=10^2$	$N=10^3$	$N=5000$	$N=10^4$	$N=10^5$	$N=10^6$
1	17.00 %	16.50 %	16.28 %	16.61 %	16.72 %	16.69 %
2	15.00 %	15.50 %	17.12 %	16.62 %	16.44 %	16.62 %
3	18.00 %	17.10 %	16.78 %	16.94 %	16.84 %	16.69 %
4	18.00 %	16.00 %	16.68 %	16.97 %	16.76 %	16.64 %
5	13.00 %	16.60 %	15.50 %	15.94 %	16.69 %	16.64 %
6	19.00 %	18.30 %	17.64 %	16.92 %	16.56 %	16.71 %

从表 13.2 可以看出，当试验的次数逐步增加时，每个点数出现的频率都向概率 $1/6\approx 16.67$ % 靠近.

例 1 中的计算机试验称为计算机**模拟试验**. 计算机模拟试验还可以解决很多其他的计算问题.

例 2　(利用几何概率估算圆周率 π)在平面上作一个边长是10 cm

的正方形，在正方形内作一个半径等于 5 cm 的圆，见图13-3.

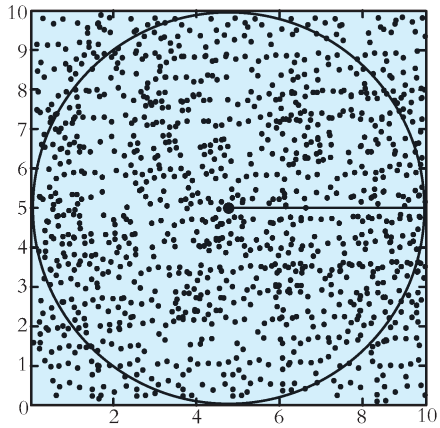


图 13-3　在 Ω 内随机投掷的1 000个质点

（1）在该正方形内随机投掷 1 个质点，计算质点落入圆 A 的概率；

（2）利用计算机模拟的方法估计 π 的值.

解　（1）试验的全集是正方形

$$\Omega=\{(x,y)\mid 0\leqslant x\leqslant 10, 0\leqslant y\leqslant 10\}.$$

事件

$$A=\{(x,y)\mid x^2+y^2\leqslant 25\}$$

是 Ω 的子集，根据圆面积的计算公式和几何概率定义 2 得到

$$P(A)=\frac{A\text{ 的面积}}{\Omega\text{ 的面积}}=\frac{25\pi}{100}=\frac{\pi}{4}.$$

（2）如果 π 是未知的，可以用如下的方法进行模拟计算.

独立重复地在 Ω 中投掷 N 个质点，对于较大的 N，质点落入圆 A 的频率

$$f_N=\frac{\text{落入 } A \text{ 的质点数}}{N}.$$

由频率和概率的关系知道 f_N 是 $P(A)$的近似，所以对较大的 N，

$$f_N\approx\frac{\pi}{4}.$$

f_N 是可以计算的，于是

$$\hat{\pi}=4f_N$$

是 π 的估计.

利用计算机在 Ω 中随机投掷 $N=10^2$，10^3，10^4，10^5 个质点，见图 13－3，把依次得到的 $\hat{\pi}$ 列入下面的表 13.3．从表 13.3 可以看出，对于较大的 N，$\hat{\pi}$ 对 π 的近似是不错的．

表 13.3

N	10^2	10^3	10^4	10^5
$\hat{\pi}$	3.080	3.148	3.160	3.149

为了看清计算机模拟结果的随机性，再次进行模拟计算时，得到的结果如下：

N	10^2	10^3	10^4	10^5
$\hat{\pi}$	3.163	3.221	3.121	3.144

例 3 A 是平面上的不规则区域，作一个长 12 m、宽 8 m 的矩形 Ω，使得 $A\subseteq\Omega$（见图 13－4）．利用计算机在 Ω 中随机投掷了 2 万个质点后，发现有 1.12 万个质点落入区域 A 中，估算 A 的面积．

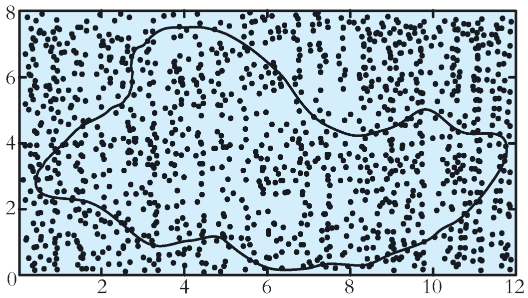


图 13－4 估算 A 的面积

解 质点落入 A 的频率是

$$f_N=\frac{1.12}{2}=0.56.$$

根据频率和概率的关系知道

$$f_N\approx\frac{A\text{ 的面积}}{\Omega\text{ 的面积}}.$$

所以

$$A\text{ 的面积}\approx f_N\times\Omega\text{ 的面积}=0.56\times12\times8=53.76\ (\text{m}^2).$$

练　习

A 是平面上的不规则区域，作一个半径为 12 cm 的圆 Ω，使得 $A \subseteq \Omega$. 在 Ω 中随机投掷了 3 000 个质点后，发现有 1 440 个质点落入区域 A 中. 估算 A 的面积.

习题 5

打乒乓球时，张三赢李四的概率是 55 %. 今天两人又要进行乒乓球比赛，为了获胜，张三应当采用以下的比赛规则（　　）.

(A) 只打 1 盘，赢者获胜

(B) 三盘两胜制，谁先领先 2 盘谁获胜

(C) 五盘三胜制，谁先领先 3 盘谁获胜

(D) 打 20 盘，这 20 盘中谁赢得多谁获胜

数学文化

概率简史

概率的概念形成于16世纪，与用投掷骰子的方法进行赌博有密切的关系.

重复投掷一枚硬币1万次，你会得到什么结果呢？如果硬币是均匀的，你会判断正面出现的频率大约是1/2吗？初看起来这是一个简单的问题，但是要在数学上证明它却不容易. 数学上首先证明这个结论的人是**伯努利**（Bernoulli），尽管他说：哪怕最笨的人，不通过别人的教诲也能理解频率大约是1/2.

伯努利1654年出生于瑞士的巴塞尔. 在他的家族成员中，程度不同地对数学的许多方面作出过贡献，其中至少有5人在概率论方面作出过贡献. 他的父亲希望他成为神职人员，但是伯努利自己更喜欢数学，他和同时代的牛顿等人保持密切的通信联系. 现在国际上的伯努利统计期刊和伯努利统计学会就是以他的名字命名的.

学习数学的人对**费马**（Fermat，1601—1665）是不陌生的. 因为“费马大定理”在前些年得到证明，费马的名声早已传播到数学的领域之外. 但是费马和概率论的关系并不为很多人所了解.

费马和**笛卡儿**（Descartes）同享发明解析几何的荣誉，但是费马最重要的研究工作是在数论方面. 费马不写论文发表，只是通过书信的形式和朋友们交流数学研究的思想和成果. 他和**帕斯卡**（Pascal，1623—1662）的通信是建立概率论的数学基础的起点.

帕斯卡出身于贵族家庭，16岁时就发表了圆锥曲线方面的数学论文. 为了帮助他父亲管理账目，他还发明了一个早期的计算机. 帕斯卡对于概率论的贡献体现在他和费马的通信中.

促使帕斯卡和费马通信的人是德梅尔，他向帕斯卡请教几个有

关赌博的问题. 1654 年 7 月 29 日帕斯卡首先给费马写信，转达了德梅尔的以下问题：投掷两个骰子 24 次，至少掷出一对 6 的概率小于 1/2. 这个概率实际上近似等于 0.491 4.

概率论的数学理论基础是由著名的前苏联数学家**柯尔莫哥洛夫**(Kolmogorov)在 1933 年建立的.

在我国，**许宝騄**教授是概率论和统计学研究的先驱，有很大的学术成就，在国际上享有盛誉，对概率论和统计学作出了杰出的贡献. 1979 年，世界著名的统计期刊《**数理统计年鉴**》(*The Annals of Statistics*) 邀请了一些著名学者撰文介绍了他的生平，高度评价了他在概率论和统计学两方面的研究工作.

多知道一点

使用计算机模拟随机试验

利用计算机和 MatLab 进行计算机模拟时，可以仿照下面程序进行.

1. 用计算机进行投掷 10^3 个硬币试验，计算正面出现的频率时，直接输入下面的语句. 括弧的内容是语句的解释，不输入.

N=10∧3；

n=unidrnd（2，1，N)－1（产生 10^3 个取值 0 或 1 的随机数，相当于投掷 10^3 个硬币）；

S=sum（n)（计算正面朝上的次数）；

M=mean（n)（计算正面出现的频率）.

2. 用计算机进行独立重复投掷一枚均匀的骰子的试验时，用下面的语句.

n=unidrnd（6，1，10∧3)（产生 10^3 个 1～6 中的随机数）；

tabulate（n)（计算出现的次数和频率）.

3. 利用计算机和几何概率估算圆周率 π 时，直接输入以下语句：

x=rand（2，10∧3)*10（在［0，10］×［0，10］中投掷 10^3 个质点）；

n=zeros（1，10∧3）；

for j=1：10∧3

if　$(x(1,j)-5)^2+(y(1,j)-5)^2<25$

n（j）=1；

end

end

M=mean（n)*4（计算 $\hat{\pi}$）

数学实验

用计算机模拟随机试验

打开用“Z＋Z 超级画板”制作的课件“圆中投豆.zjz”（可以从本书配套光盘中复制或从网站 www.zplusz.org 下载），屏幕画面如图 13－5：

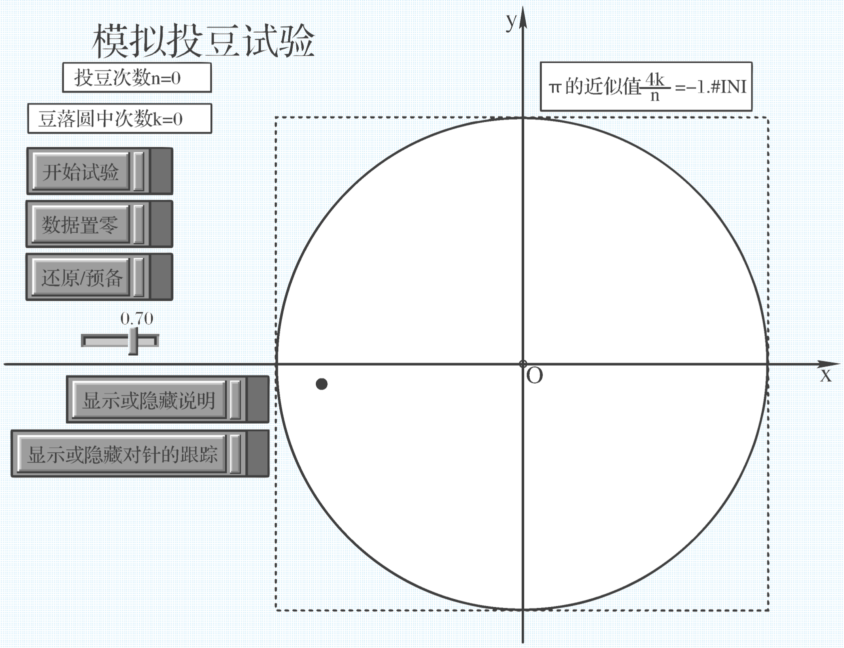


图 13－5

鼠标单击右下方“显示或隐藏说明”按钮，仔细阅读出现的文本内容；根据操作说明，单击灰色按钮上的副钮使还原；待变量尺指向 0，再单击蓝色按钮两次使数据置零；单击灰色按钮上的主钮作预备；待变量尺指向 1，单击绿色按钮开始试验.

图 13－6 是试验若干次后的画面.

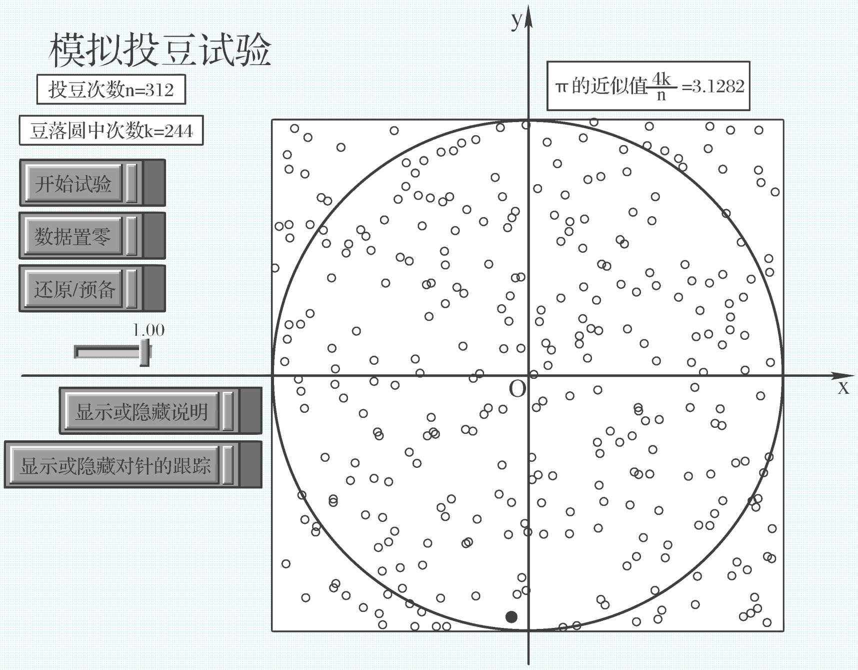


图 13－6

利用几何概率求圆周率的另一个著名的试验是投针试验．在纸上画一些等距离的平行直线，使相邻直线距离为 d．将一枚长度为 a(a<d)的针随机地投到纸上，可以算出针和直线相交的概率为

$$\mathbf{p}=\frac{2a}{\pi d}$$

想一想，若在 n 次投针中和直线相交 k 次，当 n 较大时相交频率与 π 有何关系？

打开用“Z+Z 超级画板”制作的课件“投针试验．zjz”（可以从本书配套光盘中复制或从网站 www.zplusz.org 下载），屏幕画面如图 13－7：

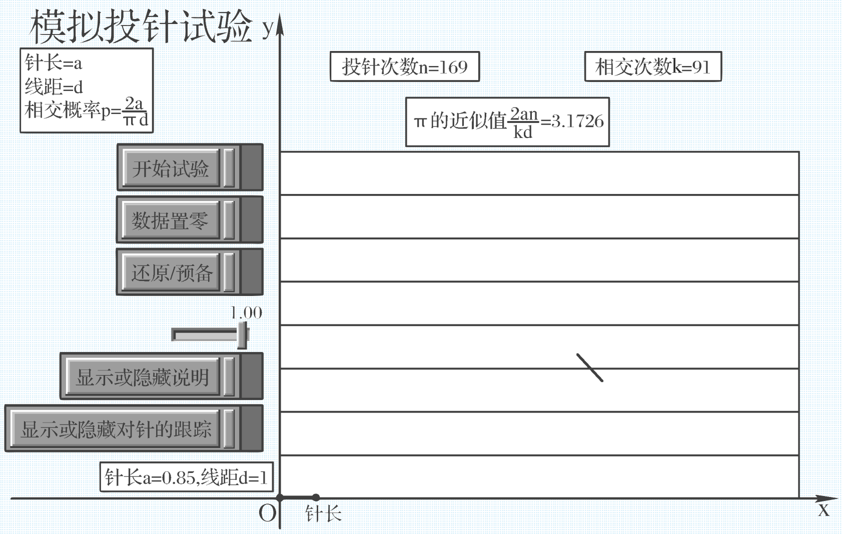


图 13－7

单击“显示或隐藏说明”按钮，仔细阅读出现的说明，依法操作，模拟投针若干次后的效果类似图 13－8：

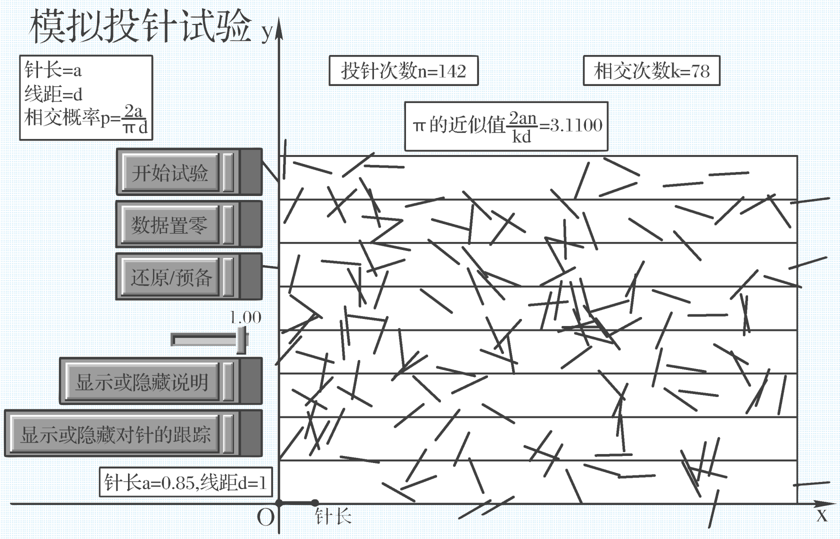


图 13－8

下面举例说明，如何在“Z＋Z 超级画板”的程序工作区直接计算模拟随机试验.

根据学过的算法知识，不难理解下面的几个程序．注意其中使

用的函数 rand(a，b)，可以每次随机地产生 a，b 之间的一个具有指定的有效数字的实数.

(1) 模拟投掷硬币输出正面向上的频率的函数 yb(m)，其中 m 是投掷次数.

键入程序：

```
yb(m) {k=m; n=0;
  while (k>0)
     {k=k-1;
      n=n+sign (rand (0, 1), 0.5);}
    n/m;}
```

执行（用 Ctrl+Enter 键，下同）后返回：

≫yb(m) #

如希望将频率表示成小数，可键入

Float(1)；

返回：

≫计算结果显示浮点数 #

要模拟试验 1 000 次，可键入

yb (1000)；

执行后返回：

≫(503)/(1000)=0.503 #

(2) 模拟掷骰子 m 次，输出 d 点向上的频率的函数 sz(m)：

```
sz(m,d){k=m; n=0;
           while(k>0)
             {k=k-1;
        if (floor(rand(1,7))==d) {n=n+1;}}
          n/m;}
```

键入程序执行后返回：

≫sz(m,d)#

要模拟 300 次投掷，求出现 2 点的频率，键入程序：

sz(300,2)；

执行后返回：

≫(47)/(300)=0.156667 #

(3) 模拟投豆试验用几何概率求圆周率的近似值的程序.

```
geo(m){
k=m; n=0;
while (k>0)
{k=k-1;
if((rand(0,10)-5)^2+(rand(0,5)-5)^2<25) {n=n+1;}}
4*n/m;}
```

键入程序执行后返回：

≫geo(m) #

要模拟投豆 1 000 次的试验，只要输入

geo(1000);

执行后返回：

≫(388)/(125)=3.104 #

当然，你在计算机上运行的结果和这里可能不同.

有了这些程序，你可以在不长的时间做大量的有关随机数的试验了.

小结与复习

1. 元素：是试验的可能结果，也称为样本点或基本事件.
2. 全集：是试验的元素的集合，常用 Ω 表示. 也称为样本空间.
3. 事件：是全集的子集.
4. 古典概型：设全集 Ω 中有 n 个元素，事件 A 包含了 m 个元素. 如果 Ω 的每个元素发生的可能性相同，就称

$$P(A)=\frac{m}{n}$$

是事件 A 的概率，称这个模型是古典概型.

5. 几何概率 1：设试验的全集 Ω 是长度为正数的区间，A 是 Ω 的子区间. 如果试验的结果随机地落在 Ω 中，则称

$$P(A)=\frac{A\text{ 的长度}}{\Omega\text{ 的长度}}$$

为事件 A 的概率.

6. 几何概率 2：设试验的全集 Ω 是面积为正数的区域，A 是 Ω 的子区域. 如果元素随机地落在 Ω 中，则称

$$P(A)=\frac{A\text{ 的面积}}{\Omega\text{ 的面积}}$$

为事件 A 的概率.

7. 概率的性质：

（1）$0\leqslant P(A)\leqslant 1$；

（2）$P(\Omega)=1$；

（3）$P(\varnothing)=0$；

（4）如果 A，B 互斥，则 $P(A\cup B)=P(A)+P(B)$；

（5）$P(A)+P(\Omega\backslash A)=1$.

8. 概率和频率：在相同的条件下，将一试验独立重复 N 次，用 f_N 表

示事件 A 在这 N 次试验中发生的频率. 当 N 增加时，f_N 将在一个固定的数值 p 附近波动，这个 p 就是事件 A 的概率 $P(A)$. f_N 是 $P(A)$的估计.

复习题十三

学而时习之

1. 口袋中有标号 1～5 的球各 1 个. 为以下的试验写出全集.

   (1) 从中任取 1 个；

   (2) 从中一次任取出 2 个.

2. 投掷一枚骰子和两枚硬币，写出全集.

3. 同时投掷一枚骰子和一枚硬币，计算概率.

   (1) 硬币是正面，骰子的点数是 3；

   (2) 硬币是正面，骰子的点数是 2 或 4.

4. 3 个同学中，每个同学作业得优的概率都是 0.5. 对于下次作业，分别用集合 A，B，C 表示以下 (1)，(2)，(3) 中的事件，计算概率.

   (1) 恰好 2 个同学得优；

   (2) 至少 1 个同学得优；

   (3) 都是优；

   (4) 计算 $B\setminus A$，$B\setminus C$，$\Omega\setminus A$ 的概率.

5. 4 个同学中的每个同学能考上某重点大学的概率都是 0.5，直接用集合 A，B，C 表示以下 (1)，(2)，(3) 中的事件，并计算概率.

   (1) 至少 1 个没考上该大学；

   (2) 至少 2 个没考上该大学；

   (3) 恰好 2 个同学考上该大学；

   (4) 说明事件 A，$A\cap C$，$\Omega\setminus A$ 的含义.

6. 盒中有标号 1～4 的白球各 1 个，标号 1～3 的红球各 1 个. 从中倒出 3 个，计算以下事件的概率.

(1) 3 个都是白球；

(2) 至少 2 个白球；

(3) 至少 1 个白球.

7. 投掷两枚骰子，计算以下事件的概率.

(1) $A=$ “两枚骰子的点数之和是 2”；

(2) $B=$ “两枚骰子的点数之和是 4”；

(3) $C=$ “两枚骰子的点数之和是 6”；

(4) 以上三个事件中，哪个概率最大，为什么？

8. 一批产品有 30 个，其中含有 3 个次品，从中随机抽取 1 个. 计算：

(1) 这个产品是次品的概率；

(2) 这个产品是正品的概率.

9. 1 万张彩票中有两个大奖.

(1) 从中抽取 1 张时，中奖的概率是多少？不中奖的概率是多少？

(2) 从中有放回地随机抽取 2 张时，中奖的概率是多少？不中奖的概率是多少？

(3) 从中无放回地随机抽取 2 张时，中奖的概率是多少？不中奖的概率是多少？

温故而知新

10. 如果全集 Ω 的事件 A，B，C，D 两两互斥，用 $A\cup B\cup C\cup D$ 表示事件 A，B，C，D 中至少有一个发生. 证明：

$P(A\cup B\cup C\cup D)=P(A)+P(B)+P(C)+P(D)$.

11. 设 Ω 是长 2 cm、宽 3 cm 的长方形，A 是以 Ω 的长为半径的圆. 如果质点等可能地落在 Ω 中，

(1) 计算质点落在 A 内的概率；

(2) 计算质点落在 A 外的概率.

12. 在区间 [1，3] 中随机地投掷两个质点，计算这两个质点都落在 [1，2] 中的概率.

13. 两人在某天的 5 时至 7 时间相互独立随机到达某地会面，先到者等候 30 min 后离去.

(1) 写出试验的全集 Ω；

(2) 用集合表示出事件 $B=$ “两人相遇”；

(3) 计算这两人能相遇的概率.

上下而求索

智者千虑，必有一失！

14. 一金融公司的主要工作是进行投资，尽管每次投资前都有缜密的投资分析，但是投资失败的概率仍保留在 5 %. 利用频率和概率的关系，说明该公司一次次相互独立的投资一定有失败的时候.

15. 某射击运动员脱靶的概率是 0.01 %，如果他独立重复射击下去，必有一次脱靶发生.（利用频率和概率的关系说明.）

愚者千虑，必有一得！

16. 张三和好友李四下棋时，赢李四的概率只有 10 %. 张三不服输，不断约李四下棋. 试说明张三总有赢棋的时候.

是赌徒就要破产！

这个赌徒破产的概率是 95 %. 如果这个赌徒的欲望增加，他破产的概率会跟着增加.

17. 一个赌徒手中有 1 000 元本金，赌博时每次输赢 1 元，输赢的概率都是 1/2. 在赌博期间，一旦输光则宣告破产. 现在该赌徒决心赢到手中有 2 万元后再停止赌博. 你认为他的目的可以达到吗？（对掷硬币的试验再次理解后给出答案.）

附 录

数学词汇中英文对照表

（按词汇所在页码的先后排序）

中文名	英 文 名	页 码
算法	algorithm	2
高斯	Gauss	59
总体	population	60
个体	individual	60
均值	mean	61
样本	sample	62
观测数据	observed data	62
样本量	sample size	62
抽样	sampling	62
估计	estimator	63
方差	variance	65
标准差	standard deviation	67
随机数	random number	72
罗斯福	Roosevelt	74
兰登	Landon	74
层权	weight	79
系统抽样方法	systematic sampling method	80
频率	frequency	82
频率分布表	frequency distribution table	83
直方图	histogram	86
柏克莱	Berkeley	88
茎叶图	stemplot	90
散点图	scatter diagram	94
胡克	Hooke	100